

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:55:16 ; Search time 803 Seconds

(without alignments)
6481.969 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttagacgttatcatcg.....agtttagcggaattgatctg 6729

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3967864

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV-xlp

-Q=/cn2_1/USFTO_spool_p/US10649433/runat_02082005_165025_1814/app_query.fasta_1.6919
-DB=A Geneseq -QFWT=fastan -SUFFIX=oligo.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CGN 1.1 977 @runat 02082005_165025_1814 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THRSADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : A Geneseq 16Dec04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004as:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	363	16.7	363	ABP51639	Abp51639 Escherich
2	360	16.5	360	Aar40926	Aar40926 Wild-type
3	360	16.5	360	Aaw58898	Aaw58898 Plasmid p
4	359	16.5	360	Aaw80734	Aaw80734 Amino aci
5	359	16.5	360	Aau79467	Aau79467 pNO23097
6	359	16.5	360	Aae15450	Aae15450 Lac repte
7	359	16.5	370	Aar40927	Aar40927 CpG deple
8	359	16.5	372	ABR99495	ABR99495 Nucleotid
9	310	14.2	799	ABr43385	ABr43385 FLIPglu f
10	309	14.2	309	ABr43374	ABr43374 E. coli g

11	309	14.2	332	4	AAU29371	Aau29371 Novel mar
12	309	14.2	332	6	ADA62006	Ada62006 E. coli g
13	306	14.0	329	2	AAR21083	Aar21083 GBP-delta
14	306	14.0	366	2	AAR21085	Aar21085 GBP-delta
15	296	13.6	1277	7	AAR52701	Aar52701 Plasmid p
16	290	13.3	1293	7	ADC22811	Adc22811 Human G p
17	290	13.3	1293	7	ADH14284	Adh14284 Vector pC
18	290	13.3	1967	6	ABB82793	Abb82793 Amino aci
19	286	13.1	286	2	AAO6551	Aay6551 Escherich
20	286	13.1	872	7	ADR6586	Adr6586 E. coli cl
21	285	13.1	360	8	ADR16185	Adr16185 pET24b pl
22	284	13.0	360	2	AAI18028	Aay18028 E. coli l
23	284	13.0	360	6	ADA62016	Ada62016 E. coli m
24	284	13.0	360	6	ADA62017	Ada62017 E. coli m
25	284	13.0	360	6	ADA62011	Ada62011 E. coli m
26	284	13.0	360	6	ADA62013	Ada62013 E. coli m
27	284	13.0	360	6	ADA62001	Ada62001 E. coli w
28	284	13.0	360	6	ADA62012	Ada62012 E. coli m
29	284	13.0	360	6	ADA62014	Ada62014 E. coli m
30	284	13.0	360	6	ADA62015	Ada62015 E. coli m
31	284	13.0	1049	5	AAU75875	Aau75875 SREBP2-La
32	283	13.0	441	5	AAU75874	Aau75874 Notchi-La
33	283	13.0	1024	5	AAU75873	Aau75873 APP-Laci
34	272	12.5	360	6	ADA62018	Ada62018 E. coli m
35	244	11.2	319	8	ADI28562	Adi28562 LacIq rep
36	240	11.0	325	8	ADO76855	Adg76855 Parachyro
37	240	11.0	1651	6	ABP70771	Abp70771 Insulin r
38	239	11.0	239	3	AAB22881	Aab22881 Enhanced
39	239	11.0	239	3	AAI79583	Aay79583 EYFP sign
40	239	11.0	239	5	ABG94443	Abg94443 Protease
41	239	11.0	239	6	AAG79830	Aag79830 Yellow fl
42	239	11.0	239	6	ADA38080	Ada38080 Aequorea
43	239	11.0	239	7	ADC18356	Adc18356 EYFP (not
44	239	11.0	239	7	ADF70401	Adf70401 Aequorea
45	239	11.0	239	7	ADG32368	Adg32368 Jellyfish

ALIGNMENTS

RESULT 1
ABP51639
ID ABP51639 standard; protein; 363 AA.
XX
AC ABP51639;
XX
XX
DT 11-OCT-2002 (first entry)
XX
DE Escherichia coli lacI repressor protein SEQ ID NO:29.
XX
KW Corynebacterium glutamicum; transcriptional regulation; gene expression;
KW amino acid biosynthesis.
XX
OS Escherichia coli.
XX
PN WO200240679-A2.
XX
PD 23-MAY-2002.
XX
PF 15-NOV-2001; 2001WO-US043096.
XX
PR 15-NOV-2000; 2000US-0248219P.
XX
XX (RAYA/) RAYAPATI P.J.
XX (CRAF/) CRAFTON C.M.

Rayapati PJ, Crafton CW;

WPI; 2002-575217/61.

N-PSDB; ABQ73199.

Novel polynucleotides from Corynebacterium glutamicum useful for inducing and regulating expression of genes, including those that are involved in amino acid biosynthesis, in bacterial cells.

XX Disclosure; Page 107-109; 112pp; English.

XX The present invention describes Corynebacterium glutamicum
transcriptional regulatory region polynucleotide sequences (I). ABQ73173
to ABQ73194 represent the C. glutamicum transcriptional regulatory
regions pta, aceA, aceB, adh, aldB, poxB, ldh, amyE, malZ, bglX, gam,
glgX, hsd, pyrR, purD, hrcA, htpX, dnaK, ctc, grpE, clpB, and narX,
respectively. (I) can be used for producing a vector. (I) can also be
used for producing a transformed Corynebacterium sp. host cell, which is
useful for producing a biosynthetic product. (I) is useful for producing
a Corynebacterium sp. host cell, which is useful for producing an amino
acid. (I) is useful for regulating and enhancing the production of a
variety of products in host cells, including amino acids such as lysine,
purine nucleotides such as inosinic acid, and heterologous polypeptides.
The present sequence represents an E. coli lacI repressor protein from
the present invention

XX Sequence 363 AA;

Alignment Scores:

Pred. No.:	1.55e-303	Length:	363
Score:	363.00	Matches:	363
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	16.65%	Indels:	0
DB:	5	Gaps:	0

US-10-649-433C-1 (1-6729) x ABP51639 (1-363)

QY 5599 GTGGTGAATGTGAACACAGTAACTGATGTCGACAGTATGCGGTGTCCTTAT 5658

DB 1 ValValAsnValLysProValThrLeuYrAspValAlaGluTyrAlaGlyValSerYr 20

QY 5659 CAGACCGTTTCCCGCGTGTGAACACAGCCAGCCAGCTTCTTCGAAACCGGGAAAAA 5718

DB 21 GlnThrValSerArgValValAsnGlnAlaSerHisValSerAlaLysThrArgGluLys 40

QY 5719 GTGGNAGCGGATGCGGAGCTGAATTACATTCACACCGGTGGCAACACTGGCG 5778

DB 41 ValGluAlaAlaMetAlaGluLeuAsnTyrIleProAsnArgValAlaGlnGlnLeuAla 60

QY 5779 GSCAACAGTGTCTGCTGATGGCGTGCACCTCCAGTCTGGCCCTGCACGCGCGTCG 5838

DB 61 GlyLysGlnSerLeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSer 80

QY 5839 CAAATTGTCGGCGATTAAATCTCCGCGCGATCAACTGGGTGCGAGCGTGGTGTGTCG 5898

DB 81 GlnIleValAlaAlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValSer 100

QY 5899 ATGTAGAACNAGCGGTGCGAGCCTGTAAGCGGCGGTGCACAACTTCTCGCGCAA 5958

DB 101 MetValGluArgSerGlyValGluAlaCysLysAlaAlaValAlaHisAsnLeuAlaGln 120

QY 5959 CGCGTCAGTGGCGTCACTAATTAACATATCCGTGGATGACAGGATCCCATTTCTGTGAA 6018

DB 121 ArgValSerGlyLeuIleAsnTyrProLeuAspAspGlnAspAlaIleAlaValGlu 140

QY 6019 GCTGCTGCACATAATTTCCGCGGTATTTCTTGATGTCTCTGACACACACCCATCAAC 6078

DB 141 AlaAlaCysThrAsnValProAlaLeuPheLeuAspValSerAspGlnThrProIleAsn 160

QY 6079 AGTATTATTTCTCCATGAGACCGGTAGCGACTGGGCGGTGGACATCTGTCGATTG 6138

DB 161 SerIleIlePheSerHisGluAspGlyThrArgLeuGlyValGluHisLeuValAlaLeu 180

QY 6139 GGTCAACAGCAATCGCGTGTAGCGGCCCATTAAGTTCTGTCCTCGCGCGTCTGGGT 6198

DB 181 GlyHisGlnGlnIleAlaLeuLeuAlaGlyProLeuSerSerValSerAlaArgLeuArg 200

QY 6199 CTGGTGGTGGCATAAATATCTACTCGCAATCAAAATTCAGCCGATAGCGGAACGGAA 6258

DB 201 LeuAlaGlyTyrHisLysTyrLeuThrArgAsnGlnIleGlnProIleAlaGluArgGlu 220

QY 6259 GCGACTGGAGTCGATGTCGGTTTCAACAAACCATGCAATGCTGAATGAGGCGATC 6318

DB 221 GlyAspTrpSerAlaMetSerGlyPheGlnGlnThrMetGlnMetLeuAsnGluGlyIle 240

QY 6319 GTTCCCACTGGAGTGGTTGGCCAAACGATCAGATGGCGCTGGCGCAATGCGCGCATT 6378

DB 241 ValProThrAlaMetLeuValAlaAsnAspGlnMetAlaLeuGlyAlaMetArgAlaIle 260

QY 6379 ACCGAGTCCGGCTGCGGTGGTGGCGATATCTCGGTAGTGGATACGATACCGAA 6438

DB 261 ThrGluSerGlyLeuArgValGlyAlaAspIleSerValValGlyTyrAspAspThrGlu 280

QY 6439 GACAGCTCATGTTATATCCCGCGCTCAACCAACCATCAACAGGATTTTCGCTGCTGGG 6498

DB 281 AspSerSerCysTyrIleProProSerThrThrIleLysGlnAspPheArgLeuLeuGly 300

QY 6499 CAAACACGCGTGGACCGCTTCTCTCAGGCGCCAGGCGGTGAAGGGCAATCAG 6558

DB 301 GlnThrSerValAspArgLeuLeuGlnLeuSerGlnGlyGlnAlaValLysGlyAsnGln 320

QY 6559 CTGTTGCGCGTCTCACTGGTGAAGAAACCAACCCCTGGCGGCCCAATACGAAACCGCC 6618

DB 321 LeuLeuProValSerLeuValLysArgLysThrThrLeuAlaProAsnThrGlnThrAla 340

QY 6619 TCTCCCGCGCGTGGCGGATTCATTAATGACGCTGGCAGCAGCAGAGTTTCCGCTGCTGAA 6678

DB 341 SerProArgAlaLeuAlaAspSerLeuMetGlnLeuAlaArgGlnValSerArgLeuGlu 360

QY 6679 AGCGGCGCAG 6687

DB 361 SerGlyGln 363

RESULT 2

AAAR40926

ID AAR40926 standard; protein; 360 AA.

XX AAR40926;

XX 25-MAR-2003 (revised)

DT 09-JAN-2003 (revised)

DT 22-FEB-1994 (first entry)

XX Wild-type lacI.

XX Bacteria; CpG doublet; vertebrate; mammalian; methylation; marker;

KW reporter; transgene; mutagen; carcinogen; mutation; assay;

KW chemical agent.

OS Bacteriophage lambda.

XX WO9317123-A1.

PD 02-SEP-1993.

XX 26-FEB-1993; 93WO-US001676.

XX 27-FEB-1992; 92US-00842664.

XX (OHIS) UNIV OHIO STATE.

XX Stambrook PJ;

XX WPI; 1993-288426/36.

DR N-PSDB; AAQ48615.

XX Modified mutagenicity test - comprises assaying transgenic animals for

PT genetic mutations, utilises native amount of CpG nucleotide(s).

XX Disclosure; Page 52-53; 79pp; English.

PS Bacterial genes exhibit a much higher frequency of occurrence of the CpG

CC doublet than do vertebrate genes. As a result, a bacterial gene

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 19:39:42 ; Search time 159.5 Seconds
(without alignments)
6298.603 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttagacgttatcatcg.....agttagcgcgaattgatctg 6729

Scoring table:

OLIGO Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Word size: 1

Total number of hits satisfying chosen parameters: 903960

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10649433/runat_02082005_165026_1844/app_query.fasta_1.6919
-DB=issued_Patents_AA -Qfmt=fastan -SUFFIX=Oligo.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649433 @CNC 1.1 178 @runat_02082005_165026_1844 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A COMB pep: *
2: /cgn2_6/ptodata/1/1aa/5B COMB pep: *
3: /cgn2_6/ptodata/1/1aa/6A COMB pep: *
4: /cgn2_6/ptodata/1/1aa/6B COMB pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS COMB pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	360	16.5	360	5	PCT-US93-01676A-2
2	359	16.5	370	5	PCT-US93-01676A-4
3	290	13.3	1293	4	US-09-170-496D-292
4	290	13.3	1293	4	US-09-364-425B-57
5	286	13.1	286	4	US-09-555-510B-9
6	286	13.1	286	4	US-10-231-013-9
7	285	13.1	360	3	US-09-025-769B-293
8	285	13.1	360	4	US-09-490-070A-293
9	285	13.1	360	4	US-09-490-153-293
10	285	13.1	360	4	US-09-490-324-293
11	284	13.0	1049	3	US-09-522-666-6
12	283	13.0	441	3	US-09-522-666-4

13	283	13.0	1024	3	US-09-522-666-2	Sequence 2, Appli
14	244	11.2	319	3	US-09-044-796A-13	Sequence 13, Appl
15	244	11.2	319	4	US-09-725-460A-13	Sequence 13, Appl
16	239	11.0	239	4	US-09-513-783A-44	Sequence 44, Appl
17	239	11.0	239	4	US-09-430-656-44	Sequence 44, Appl
18	239	11.0	541	4	US-09-513-783A-34	Sequence 34, Appl
19	239	11.0	541	4	US-09-430-656-34	Sequence 34, Appl
20	239	11.0	812	4	US-09-513-783A-4	Sequence 4, Appli
21	239	11.0	812	4	US-09-513-783A-6	Sequence 6, Appli
22	239	11.0	812	4	US-09-430-656-4	Sequence 6, Appli
23	239	11.0	812	4	US-09-430-656-6	Sequence 6, Appli
24	239	11.0	1610	4	US-09-513-783A-22	Sequence 22, Appl
25	239	11.0	1610	4	US-09-430-656-22	Sequence 22, Appl
26	238	10.9	1056	4	US-09-513-783A-32	Sequence 32, Appl
27	238	10.9	1056	4	US-09-430-656-32	Sequence 32, Appl
28	204	9.4	286	1	US-07-721-775A-2	Sequence 2, Appli
29	204	9.4	286	1	US-08-339-658-2	Sequence 2, Appli
30	204	9.4	286	3	US-09-263-933-7	Sequence 7, Appli
31	204	9.4	286	3	US-09-263-933-14	Sequence 14, Appl
32	204	9.4	286	3	US-09-263-933-21	Sequence 21, Appl
33	204	9.4	286	3	US-09-025-769B-265	Sequence 265, App
34	204	9.4	286	3	US-09-025-769B-362	Sequence 362, App
35	204	9.4	286	4	US-09-919-901-7	Sequence 7, Appli
36	204	9.4	286	4	US-09-919-901-14	Sequence 14, Appl
37	204	9.4	286	4	US-09-919-901-21	Sequence 21, Appl
38	204	9.4	286	4	US-09-490-070A-265	Sequence 265, App
39	204	9.4	286	4	US-09-490-070A-362	Sequence 362, App
40	204	9.4	286	4	US-09-490-153-265	Sequence 265, App
41	204	9.4	286	4	US-09-490-153-362	Sequence 362, App
42	204	9.4	286	4	US-10-191-966-7	Sequence 7, Appli
43	204	9.4	286	4	US-10-191-966-14	Sequence 14, Appl
44	204	9.4	286	4	US-10-191-966-21	Sequence 21, Appl
45	204	9.4	286	4	US-09-490-324-265	Sequence 265, App

ALIGNMENTS

RESULT 1
PCT-US93-01676A-2
; Sequence 2, Application PC/TUS9301676A
; GENERAL INFORMATION:
; APPLICANT: Scambrook, Peter J.
; TITLE OF INVENTION: MUTAGENICITY TESTING USING REPORTER
; GENES WITH MODIFIED METHYLATION FREQUENCIES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01676A
; FILING DATE: 19930226
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/842,644
; FILING DATE: 02-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: STAMBOOK 1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-01676A-2

Alignment Scores:

Pred. No.: 4,566-298 Length: 360
Score: 360.00 Matches: 360
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.51% Indels: 0
DB: 5 Gaps: 0

US-10-649-433C-1 (1-6729) x PCT-US93-01676A-2 (1-360)

QY 5608 GTGAACACCAAGTATACGATGTCGACGATGTCGGGTGCTCTTATCAGACCGGT 5667
Db 1 ValLysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrVal 20
QY 5668 TCCCGCGTGTGAACACGACGACGACGCTTCTGCGAAACGCGGAAAAAGTGAACGG 5727
Db 21 SerArgValValAsnGlnAlaSerHisValSerAlaLysThrArgGluLysValGluAla 40
QY 5728 GCGATGGCGGAGCTGAATTACATTCCTCCCAACGCGTGGCACAACACTGCGCGGCAACAG 5787
Db 41 AlaMetAlaGluLeuAsnTyrIleProAsnArgValAlaGlnGlnLeuAlaGlyLysGln 60
QY 5788 TCGTTGCTGATGGCGTGGCCACCTCCAGTCTGGCCCTGCACGCGCCGCGCAAAATGTC 5847
Db 61 SerLeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSerGlnIleVal 80
QY 5848 GCGCGGATTAATCTCGCGCGCATCAACTGGGTGCCAGCGTGGTGTCTCGATGTAGAA 5907
Db 81 AlaAlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValValSerMetValGlu 100
QY 5908 CGAAGCGCGTGGAGCTGTAAAGCGGCGGTGCACATCTCTCGCGCAACGCTCAGT 5967
Db 101 ArgSerGlyValGluAlaCysLysAlaAlaValHisAsnLeuLeuAlaGlnArgValSer 120
QY 5968 GGGCTGATCAATTAATCCGCTGATGACACGAGATGCTTGTGTGGAAGTGCCTGCG 6027
Db 121 GlyLeuIleIleAsnTyrProLeuAspAspGlnAspAlaIleAlaValGluAlaCys 140
QY 6028 ACTAATGTCGCGCGTTATTTCTGATGCTCTGACACGACACCATCAACAGATTATT 6087
Db 141 ThrAsnValProAlaLeuPheLeuAspValSerAspGlnThrProIleAsnSerIleIle 160
QY 6088 TTCTCCCATGAAGCGGTACGCGTGGCGTGGAGCATCTGTCGCATTCGGTCCACGAG 6147
Db 161 PheSerHisGluAspGlyThrArgLeuGlyValGluHisLeuValAlaLeuGlyHisGln 180
QY 6148 CAAATCGCGCTGTAGCGGCGCCATTAAGTTCTGTCTCGCGCGTCTGCTGCTGCGC 6207
Db 181 GlnIleAlaLeuLeuAlaGlyProLeuSerSerValSerAlaArgLeuArgLeuAlaGly 200
QY 6208 TGGCATAAATCTCACTCGCAATCAAAATTCAGCCGATAGCGGAACGGGAAGCGACTGG 6267
Db 201 TrpHisLysTyrLeuThrArgAsnGlnIleGlnProIleAlaGluArgGluGlyAspTrp 220
QY 6268 AGTGCATGTCGGGTTTTCACAAACCATGCAATGCTGAATGAGGGCATCGTTCCCACT 6327
Db 221 SerAlaMetSerGlyPheGlnThrMetGlnMetLeuAsnGluGlyIleValProThr 240
QY 6328 GCGATGCTGGTTCACAGATGCGATGCGTGGCGCAATCGCGCCATTACCGAGTCC 6387
Db 241 AlaMetLeuValAlaAsnAspGlnMetAlaLeuGlyAlaMetArgAlaIleThrGluSer 260
QY 6388 GGGCTGCGGTTGGTGGGATATCTCGGTAGTGGGATACGAGATACCGAAGACAGCTCA 6447
Db 261 GlyLeuArgValGlyAlaAspIleSerValValGlyTyrAspAspThrGluAspSerSer 280

QY 6448 TGTATATCCCGCGCTCAACACCATCAACAGAGATTTTCCGCTGCTGGGCAACACGAGC 6507
Db 281 CysTyrIleProProSerThrThrIleLysGlnAspPheArgLeuLeuGlyGlnThrSer 300
QY 6508 GTGACCGCTTGTGCTCAACTCTCTCAGGCGCAGCGGTGAAGGGCAATCAGCTGTGCC 6567
Db 301 ValAspArgLeuLeuGlnLeuSerGlnGlyGlnAlaValLysGlyAsnGlnLeuLeuPro 320
QY 6568 GTCTCAGTGTGAAAGAAAAACACCTCTGCGGCGCCCAATACGCAACCGCTCTCCCGC 6627
Db 321 ValSerLeuValLysArgLysThrThrLeuAlaProAsnThrGlnThrAlaSerProArg 340
QY 6628 GCGTTGGCCGATTCAATTAATCAGCTGCGCAGCAGAGTTTCCCGACTCGAAAGCGGCGAG 6687
Db 341 AlaLeuAlaAspSerLeuMetGlnLeuAlaArgGlnValSerArgLeuGluSerGlyGln 360

RESULT 2

PCT-US93-01676A-4

; Sequence 4, Application PC/TUS9301676A

; GENERAL INFORMATION:

; APPLICANT: Stambrook, Peter J.

; TITLE OF INVENTION: MUTAGENICITY TESTING USING REPORTER

; NUMBER OF SEQUENCES: 8 GENES WITH MODIFIED METHYLATION FREQUENCIES

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Browdy and Neimark

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/01676A

; FILING DATE: 19930226

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/842,644

; FILING DATE: 02-FEB-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Cooper, Iver P.

; REGISTRATION NUMBER: 28,005

; REFERENCE/DOCKET NUMBER: STAMBROOK 1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 370 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US93-01676A-4

Alignment Scores:

Pred. No.: 3,23e-297 Length: 370
Score: 359.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.47% Indels: 0
DB: 5 Gaps: 0

US-10-649-433C-1 (1-6729) x PCT-US93-01676A-4 (1-370)

QY 5611 AAACAGTAAGTTATACGATGTCGACGATGTCGGGTGCTCTTATCAGACCGTTCC 5670
Db 2 LysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrValSer 21

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 20:07:07 ; Search time 801 Seconds
(without alignments)

6545.998 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttagacgcttatcatcg.....agttagcgcgaattgatctg 6729

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1745140 seqs, 389608008 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3345576

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10649433/runat_02082005_165027_1863/app_query.fasta_1.6919
-DB=Published Applications AA -QWITS=fastan -SUFFIX=Oligo.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649433 @CIGN 1.1 964 @runat_02082005_165027_1863
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=60 -XGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pdb.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pdb.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pdb.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pdb.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pdb.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pdb.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pdb.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pdb.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pdb.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pdb.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pdb.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pdb.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pdb.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pdb.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pdb.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pdb.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_NEW_PUB.pdb.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pdb.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pdb.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pdb.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pdb.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	363	16.7	363	10	US-09-987-763-29 Sequence 29, Appl
2	359	16.5	372	16	US-10-469-881-3 Sequence 3, Appl
3	309	14.2	309	16	US-10-840-796-1 Sequence 1, Appl
4	309	14.2	332	13	US-10-197-053-6 Sequence 6, Appl
5	309	14.2	332	17	US-10-893-671-82 Sequence 82, Appl
6	290	13.3	1293	14	US-10-251-385-292 Sequence 292, Appl
7	290	13.3	1293	18	US-10-668-035-57 Sequence 57, Appl
8	290	13.3	1967	16	US-10-477-044-1 Sequence 1, Appl
9	286	13.1	286	14	US-10-231-013-9 Sequence 9, Appl
10	286	13.1	286	16	US-10-842-534-9 Sequence 9, Appl
11	284	13.0	360	13	US-10-197-053-1 Sequence 1, Appl
12	244	11.2	319	9	US-09-725-460A-13 Sequence 13, Appl
13	244	11.2	319	20	US-11-004-853-15 Sequence 15, Appl
14	239	11.0	239	14	US-10-162-593-3 Sequence 3, Appl
15	239	11.0	239	14	US-10-100-957A-44 Sequence 44, Appl
16	239	11.0	239	15	US-10-370-570-5 Sequence 5, Appl
17	239	11.0	239	15	US-10-442-148A-5 Sequence 5, Appl
18	239	11.0	239	16	US-10-724-178-4 Sequence 4, Appl
19	239	11.0	239	17	US-10-763-976A-42 Sequence 42, Appl
20	239	11.0	239	17	US-10-763-976A-55 Sequence 55, Appl
21	239	11.0	239	17	US-10-505-486-24 Sequence 24, Appl
22	239	11.0	253	15	US-10-442-148A-6 Sequence 6, Appl
23	239	11.0	541	14	US-10-100-957A-34 Sequence 34, Appl
24	239	11.0	643	16	US-10-838-770-4 Sequence 4, Appl
25	239	11.0	812	14	US-10-100-957A-4 Sequence 4, Appl
26	239	11.0	812	14	US-10-100-957A-6 Sequence 6, Appl
27	239	11.0	1610	14	US-10-100-957A-22 Sequence 22, Appl
28	238	10.9	1056	14	US-10-100-957A-32 Sequence 32, Appl
29	237	10.9	238	17	US-10-505-486-5 Sequence 5, Appl
30	237	10.9	387	11	US-09-927-876-87 Sequence 87, Appl
31	237	10.9	387	15	US-10-360-149-87 Sequence 87, Appl
32	237	10.9	387	15	US-10-457-047-87 Sequence 87, Appl
33	237	10.9	387	17	US-10-811-081-87 Sequence 87, Appl
34	237	10.9	387	17	US-10-913-944-87 Sequence 87, Appl
35	237	10.9	545	14	US-10-214-932-52 Sequence 52, Appl
36	237	10.9	548	14	US-10-214-932-76 Sequence 76, Appl
37	237	10.9	1192	14	US-10-214-932-54 Sequence 54, Appl
38	225	10.3	360	13	US-10-137-053-2 Sequence 2, Appl
39	204	9.4	263	15	US-10-668-778-2 Sequence 2, Appl
40	204	9.4	264	17	US-10-877-952-18 Sequence 18, Appl
41	204	9.4	286	10	US-09-919-901-7 Sequence 7, Appl
42	204	9.4	286	10	US-09-919-901-14 Sequence 14, Appl
43	204	9.4	286	10	US-09-919-901-21 Sequence 21, Appl
44	204	9.4	286	11	US-09-837-306-354 Sequence 354, App
45	204	9.4	286	14	US-10-191-966-7 Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-09-987-763-29
; Sequence 29, Application US/09987763
; Publication No. US20030017553A1
; GENERAL INFORMATION:
; APPLICANT: Rayapati, P. John
; APPLICANT: Crafton, Corey M.
; TITLE OF INVENTION: Nucleotide Sequences for Transcriptional Regulation in
; TITLE OF INVENTION: Corynebacterium glutamicum
; FILE REFERENCE: 1533.1940002
; CURRENT APPLICATION NUMBER: US/09/987,763
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/248,219
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Escherichia coli

US-09-987-763-29

Alignment Scores:

Pred. No.: 0 Length: 363
Score: 363.00 Matches: 363
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.65% Indels: 0
DB: 10 Gaps: 0

US-10-649-433C-1 (1-6729) x US-09-987-763-29 (1-363)

QY 5599 GTGGTGAATGTCARACAGTACGTTATACGATGTCGACAGTATGCGGTGTCCTTAT 5658
Db 1 ValValAsnValLysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyr 20
QY 5659 CAGACCGTTTCCCGCGTGGTGAACACAGCCAGCCACCGTTTCTCGCAAAACCGCGGAAAAA 5718
Db 21 GlnThrValSerArgValValAsnGlnAlaSerHisValSerAlaLysThrArgGluLys 40
QY 5719 GTGGAAGCGCGATGCGGAGCTGAATTACATCCCAACCGCGTGGCCACAACAATCGCG 5778
Db 41 ValGluAlaAlaMetAlaGluLeuAsnTyrIleProAsnArgValAlaGlnGlnLeuAla 60
QY 5779 GGCAAACAGTCGTTGCTGATTGGCGTTCGCCACTCCAGTCTGGCCCTGCACGCGCGTCG 5838
Db 61 GlyLysGlnSerLeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSer 80
QY 5839 CAAATTGTCGCGCGATTAAATCTCGCGCGATCAACTGGGTGCGCAGCGTGGTGTGTCG 5898
Db 81 GlnIleValAlaAlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValSer 100
QY 5899 ATGGTAGAACGAGCGCGTGAAGCCTGTAAAGCGCGGTGCACAATCTTCTCGCGCAA 5958
Db 101 MetValGluArgSerGlyValGluAlaCysLysAlaAlaValHisAsnLeuLeuAlaGln 120
QY 5959 CGCGTCAGTGGCTCATTAATCAATATCCGCTGGATGACAGGATGCCATTGCTGTGAA 6018
Db 121 ArgValSerGlyLeuIleLeuAsnTyrProLeuAspGlnAspAlaIleAlaValGlu 140
QY 6019 GTCGCTGCACATAATGTTCCGCGGTTATTTCTGTATGTCTCTGACGACACCCATCAAC 6078
Db 141 AlaAlaCysThrAsnValProAlaLeuPheLeuAspValSerAspGlnThrProIleAsn 160
QY 6079 AGTATTATTTCTCCATGAACACGCTAGCGACTGGCGGTGGACCATCTGCTCCATTG 6138
Db 161 SerIleIlePheSerHisGluAspGlyThrArgLeuGlyValGluHisLeuValAlaLeu 180
QY 6139 GGTACACAGCAATCGCGTGTAGCGGCCCATTAAGTTCTGCTCGCGCGTCTGCGT 6198
Db 181 GlyHisGlnGlnIleAlaLeuLeuAlaGlyProLeuSerSerValSerAlaArgLeuArg 200
QY 6199 CTGGCTGGCTGGCATAAATATCTACTCGCAATCAAATTCAGCCGATAGCGGAACGGAA 6258
Db 201 LeuAlaGlyTyrHisLysTyrLeuThrArgAsnGlnIleGlnProIleAlaGluArgGlu 220
QY 6259 GCGAGTCGAGTGCATGCTCGGTTTCAAACCAACCATGCMAATCTGAATGAGGCGATC 6318
Db 221 GlyAspTyrSerAlaMetSerGlyPheGlnGlnThrMetGlnMetLeuAsnGluGlyIle 240
QY 6319 GTTCCACTCGCATGCTGGTTCGCCAACATCAGATGGCGCTGGGCGCAATGCGCCCAAT 6378
Db 241 ValProThrAlaMetLeuValAlaAsnAspGlnMetAlaLeuGlyAlaMetArgAlaIle 260
QY 6379 ACCGAGTCGCGGCTCGGCTGGTGGGATATCTCGGTAGTGGGATACGACGATACCGAA 6438
Db 261 ThrGluSerGlyLeuArgValGlyAlaAspIleSerValValGlyTyrAspAspThrGlu 280
QY 6439 GACAGTCATGTATATCCGCGGTCAACCCACCATCAACAGGATTTTCGGCTGCTGGG 6498
Db 281 AspSerSerCysTyrIleProProSerThrThrIleLysGlnAspPheArgLeuLeuGly 300
QY 6499 CAAACACAGCGTGGACCGCTTGTCGAATCTCTCAGGGCCAGCGCGGTGAAGGGCAATCAG 6558

Db 301 GlnThrSerValAspArgLeuLeuGlnLeuSerGlnGlyGlnAlaValLysGlyAsnGln 320
QY 6559 CTGTTGCCGCTCTCACTGGTCAAAAAGAAACACCCCTGGCGCCCAATACGCAAAACCGCC 6618
Db 321 LeuLeuProValSerLeuValLysArgLysThrThrLeuAlaProAsnThrGlnThrAla 340
QY 6619 TCTCCCGCGCGTGGCGCGATTCAATTATGAGCTGGGACGACAGGTTTCCGACTGGA 6678
Db 341 SerProArgAlaLeuAlaAspSerLeuMetGlnLeuAlaArgGlnValSerArgLeuGlu 360
QY 6679 AGCGGGCAG 6687
Db 361 SerGlyGln 363
RESULT 2
US-10-469-881-3
; Sequence 3, Application US/10469881
; Publication No. US20040171824A1
; GENERAL INFORMATION:
; APPLICANT: The University of Virginia Patent Foundation
; APPLICANT: Scrable, Heidi J
; APPLICANT: Cronin, Carolyn A
; TITLE OF INVENTION: A Lac Operator-Repressor System
; FILE REFERENCE: 00663-02
; CURRENT APPLICATION NUMBER: US/10469,881
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60/281,322
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/273,480
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-469-881-3
Alignment Scores:
Pred. No.: 0 Length: 372
Score: 359.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.47% Indels: 0
DB: 16 Gaps: 0
US-10-649-433C-1 (1-6729) x US-10-469-881-3 (1-372)
QY 5611 AAACCACTAAGCTTATACGATGTCGACAGTATGCCGTGTCTTATCAGACCGTTTCC 5670
Db 2 LysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrValSer 21
QY 5671 CGCGTGGTGAACACGAGCCAGCCACGTTTCTCGCAAAACCGGAAAAAGTGAAGCGCG 5730
Db 22 ArgValValAsnGlnAlaSerHisValSerAlaLysThrArgGluLysValGluAla 41
QY 5731 ATGCGGAGCTGAATTAATTCCTCCCAACCGCGTGGCACAACAACCTGCGCGGCGCAACGTCG 5790
Db 42 MetAlaGluLeuAsnTyrIleProAsnArgValAlaGlnGlnLeuAlaGlyLysGlnSer 61
QY 5791 TTGCTGATTGGGTGGCCACCTCCAGTCTGCCCGCTGCACGCGCGCTCCAAATGTCGCG 5850
Db 62 LeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSerGlnIleValAla 81
QY 5851 GCGATTAAATCTCGCGCGATCAACTGGGTGCCAGCGTGGTGGTTCGATGCTAGAACGA 5910
Db 82 AlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValValSerMetValGluArg 101
QY 5911 AGCGCGTGAAGCCTGTAAAGCGCGGTGACAACTCTTCTCGCGCAACGCGTCAAGTGG 5970
Db 102 SerGlyValGluAlaCysLysAlaAlaValHisAsnLeuLeuAlaGlnArgValSerGly 121

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 19:06:10 ; Search time 27848 Seconds
(without alignments)
11708.387 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 6729
Sequence: 1 gtttgacagcttatcatcga.....agttagcgcgaattgatctg 6729

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sta.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3855	57.3	4806	12 AF050464	AF050464 Expressio
2	3804	56.5	5013	6 AR264513	AR264513 Sequence
3	3753	55.8	4176	6 A13038	A13038 Artificial
4	3753	55.8	4176	12 SYNTRC99A	M22744 Cloning vec
5	3753	55.8	4239	6 A29289	A29289 pSEC-Bp1 co
6	3753	55.8	5271	12 CVU089936	U89936 Cloning vec
7	3752	55.8	4476	6 AR102989	AR102989 Sequence
8	3752	55.8	4476	6 AR172783	AR172783 Sequence
9	3703	55.0	4214	6 BD273391	BD273391 P450/acet
10	3703	55.0	4214	6 AX027488	AX027488 Sequence
11	3702	55.0	4176	12 XXU13872	U13872 pirc99a clo
12	3702	55.0	4203	6 AX576404	AX576404 Sequence
13	3702	55.0	4204	6 AX576405	AX576405 Sequence
14	3702	55.0	4205	6 AX576406	AX576406 Sequence
15	3702	55.0	4216	6 AX576407	AX576407 Sequence
16	3702	55.0	4217	6 AX576408	AX576408 Sequence
17	3702	55.0	4218	6 AX576409	AX576409 Sequence
18	3384	50.3	4232	12 SYNLACIQ	L24193 Cloning vec
19	3384	50.3	4735	12 SYNBTALAC	L24367 Cloning vec

20	2729	40.6	4357	6 AR431490	AR431490 Sequence
21	2729	40.6	4357	6 AX072889	AX072889 Sequence
22	2691	40.0	4770	12 AF339022	AF339022 Expressio
23	2676	39.8	6862	6 A10361	A10361 Artificial
24	2661	39.5	6886	6 AX572872	AX572872 Sequence
25	2661	39.5	6886	6 AX573224	AX573224 Sequence
26	2658	39.5	5797	6 CQ857478	CQ857478 Sequence
27	2658	39.5	5797	6 AX786864	AX786864 Sequence
28	2648	39.4	5021	6 BD211261	BD211261 Phagocyt
29	2648	39.4	5021	6 AX008866	AX008866 Sequence
30	2648	39.4	7268	6 BD187081	BD187081 Polypepti
31	2648	39.4	7325	6 BD187078	BD187078 Polypepti
32	2648	39.4	7439	6 BD187077	BD187077 Polypepti
33	2645	39.3	5885	1 STAFRA	L42764 Staphylococ
34	2613	38.8	6806	6 AX378208	AX378208 Sequence
35	2612	38.8	4593	12 XXU02439	U02439 Cloning vec
36	2612	38.8	5075	12 U02444	U02444 Cloning vec
37	2527	37.6	4593	6 AR102988	AR102988 Sequence
38	2527	37.6	4593	6 AR172782	AR172782 Sequence
39	2527	37.6	4593	12 CVPKX2332	X70478 Plasmid pKK
40	2526	37.5	5289	12 AF121783	AF121783 Expressio
41	2526	37.5	7050	12 AF121785	AF121785 Expressio
42	2491	37.0	4921	6 AX768201	AX768201 Sequence
43	2491	37.0	4935	6 AX768200	AX768200 Sequence
44	2491	37.0	4945	6 AX768202	AX768202 Sequence
45	2491	37.0	4951	6 AX768203	AX768203 Sequence

ALIGNMENTS

RESULT 1
AF050464
LOCUS AF050464 4806 bp DNA circular SYN 19-NOV-1998
DEFINITION Expression vector pKIL-HIS3, complete sequence.
ACCESSION AF050464
VERSION AF050464.1 GI:3892175
KEYWORDS
SOURCE Expression vector pKIL-HIS3
ORGANISM Expression vector pKIL-HIS3
REFERENCE 1 (bases 1 to 4806)
AUTHORS Van Reeth, T., Dreze, P.L., Szpirer, J., Szpirer, C. and Gabant, P.
TITLE Positive selection vectors to generate fused genes for the expression of his-tagged proteins
JOURNAL Biochimie 25 (5), 898-904 (1998)
MEDLINE 99039034
REFERENCE 2 (bases 1 to 4806)
AUTHORS Van Reeth, T., Gabant, P., Dreze, P.L., Szpirer, J. and Szpirer, C.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1998) Departement de Biologie Molculaire, Universite Libre de Bruxelles, 67, rue des Chevaux, Rhode St. Genese 1640, Belgium
FEATURES
source
1. 4806
/organism="Expression vector pKIL-HIS3"
/mol_type="genomic DNA"
/specific_host="Escherichia coli"
/db_xref="taxon:74458"
28..1002
/gene="ccdB"
28..58
/gene="ccdB"
250..774
/gene="ccdB"
/note="CcdB fusion protein"
/codon_start=1
/transl_table=11
/product="CcdB protein"
/protein_id="AAC78290.1"
/db_xref="GI:3892176"
/translation="MSRRRRQAFSSSSSLIEGSDLVAVLQRDWNPGVTQLRRL"

AAHPPFASWRNSEARTRPSQOLBSLYVRQFKYTYKRSRYRLFDVDVQSDIIDTPG
RMV1PLASARLLSDKVSRELYPVVHHGDESRRMTTDMASVPVSVIGEEVADIUSHRE
NDIKNAINLMFWGI"

misc_feature

/gene="ccdB"
/note="unique NruI (blunt) cloning site that can be used
for generation of C-terminal His-tagged protein"

terminator

876..1002

gene

/gene="ccdB"
1313..2173
/gene="amp^r"
/note="ampicillin resistance gene"

CDS

1313..2173
/gene="amp^r"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AAC78291.1"
/db_xref="GI:3892177"

/translation="MSIQHFRVALIPFFAFLPFAHPETLVKVKDAEDQLGARVGY
IELDLSKILESFRPEERFPMWSFKVLLCGAVLSRVDAQEQQLGRIHYSQNDLVE
YSPVTEKLTIDGMVRELCSAAITMSDNTAANLLITGGPKELTAFIHNMGDVTPL
RWPELNEATPNDKRDTPMPVAMATTLTKLTGELLTLASRQQLIDWMEADKVGPL
LRSLPAGWFIADKSGAGERSGRIIAALGPDGKPSRVIVYITGTSQATMDERNQIA
EIGASLIKHW"
2319..2992

rep_origin

3645..4604

gene

/gene="lacIq"

CDS

3645..4604
/gene="lacIq"
/function="repressor of lac promoter"
/codon_start=1
/transl_table=11
/product="LacIq"
/protein_id="LAC78292.1"
/db_xref="GI:3892178"
/translation="MAELNYIPNRVAQQLAGKQSLIGVATSSIALHAPSQIVAAIKS
RADQIGASVVMVSVERGEAKAAVHNLQARVSGLLIINYPLODDDAIAVEAACTNV
PALFDVSDQTPINSIIIFSHEDGRLGVHELVALLAGLPLSVSARLRLAGW
HKYLTNRQIQIAEREGDSWMSGFQOTMOMNEGIVPTAMLVANDOMALGAMRAITE
SGLRVGADISVVGVDTRDSCYIPPSSTTIKQDFRLGQTSVDRLLQLSQGQAVKGNQ
LLPVSLVKRKTIAPTQTASPRALADSLMQLARQVSRLESSQ"

ORIGIN

Query Match 57.3%; Score 3855; DB 12; Length 4806;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2875	AGCTTGGCTGTTTCGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAAC	2934
DB	789	AGCTTGGCTGTTTCGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAAC	848
QY	2935	GCAGAAGCGGTCTGATAAAACAGAAATTTGCCTGGCGGAGTAGCGCGGTGGTCCACCTG	2994
DB	849	GCAGAAGCGGTCTGATAAAACAGAAATTTGCCTGGCGGAGTAGCGCGGTGGTCCACCTG	908
QY	2995	ACCCCATGCCGAATCTCAGAAAGTAAACCGCGTAGCGCGGATGGTAGTGTGGGTCTCCCC	3054
DB	909	ACCCCATGCCGAATCTCAGAAAGTAAACCGCGTAGCGCGGATGGTAGTGTGGGTCTCCCC	968
QY	3055	ATGCGAGAGTAGGAACCTGCCAGGCATCAATAAACAAGAAAGGCTCAGTCGAAAGACTGG	3114
DB	969	ATGCGAGAGTAGGAACCTGCCAGGCATCAATAAACAAGAAAGGCTCAGTCGAAAGACTGG	1028
QY	3115	GCCTTTTCGTTTATCTGTTGTTGTCGCTGAACCGCTCTCCTCAGTAGGACAAATCCGCGC	3174
DB	1029	GCCTTTTCGTTTATCTGTTGTTGTCGCTGAACCGCTCTCCTCAGTAGGACAAATCCGCGC	1088
QY	3175	GGAGCGGATTTGAACTGTCGAAGCAACGCGCCCGGAGGTTGCGGGAGGAGCGCCCGCA	3234
DB	1089	GGAGCGGATTTGAACTGTCGAAGCAACGCGCCCGGAGGTTGCGGGAGGAGCGCCCGCA	1148
QY	3235	TAAACTGCCAGGCATCAAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTCGCTT	3294

DB	1149	TAAACTGCCAGGCATCAAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTCGCTT	1208
QY	3295	CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTTCAAATATGTTATCCGCTCATGAGAC	3354
DB	1209	CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTTCAAATATGTTATCCGCTCATGAGAC	1268
QY	3355	AATAACCCCTGATAAATGCTTCAATAATTAAGAAAGAGAGTATGATGATTTCAACATT	3414
DB	1269	AATAACCCCTGATAAATGCTTCAATAATTAAGAAAGAGAGTATGATGATTTCAACATT	1328
QY	3415	TCCGTGTGCGCCTTATTTCCCTTTTTCGGGCATTTTTCCTCTCTCTCTCTCTCTCTCTCT	3474
DB	1329	TCCGTGTGCGCCTTATTTCCCTTTTTCGGGCATTTTTCCTCTCTCTCTCTCTCTCTCTCT	1388
QY	3475	AAACGCTGGTGAAGATGCTGAAGATCAGTTGGGTGGCAGAGTGGGTTCATCATCG	3534
DB	1389	AAACGCTGGTGAAGATGCTGAAGATCAGTTGGGTGGCAGAGTGGGTTCATCATCG	1448
QY	3535	AACTGGATCTCAACAGCGGTAAAGATCTCTTGAGAGTTTTCGCCCGGAGAAAGCTTTTCCAA	3594
DB	1449	AACTGGATCTCAACAGCGGTAAAGATCTCTTGAGAGTTTTCGCCCGGAGAAAGCTTTTCCAA	1508
QY	3595	TGATGAGCACCTTTTAAAGTCTGCTATGTCGCGGTATTTATCCCGTCTTGACCGCGGC	3654
DB	1509	TGATGAGCACCTTTTAAAGTCTGCTATGTCGCGGTATTTATCCCGTCTTGACCGCGGC	1568
QY	3655	AAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAG	3714
DB	1569	AAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAG	1628
QY	3715	TCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATTAATGTCAGTCTGTCATAA	3774
DB	1629	TCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATTAATGTCAGTCTGTCATAA	1688
QY	3775	CCATGAGTGAACACATGCGGCCCACTTACTTCTGACAAACCATCGGAGACCGAAGGAGC	3834
DB	1689	CCATGAGTGAACACATGCGGCCCACTTACTTCTGACAAACCATCGGAGACCGAAGGAGC	1748
QY	3835	TAAACCGCTTTTTCGCAACATGCGGGATCATGTAACCTGCGCTTTCGTTGGGAAACCGG	3894
DB	1749	TAAACCGCTTTTTCGCAACATGCGGGATCATGTAACCTGCGCTTTCGTTGGGAAACCGG	1808
QY	3895	AGCTGAATGAAGCCATACCAACGAGCGGTGACACACCATGCTGTAGCAATGGCAA	3954
DB	1809	AGCTGAATGAAGCCATACCAACGAGCGGTGACACACCATGCTGTAGCAATGGCAA	1868
QY	3955	CAACGTTGCGCAAACTATTAACTGCGGAACTACTTACTTCTAGCTTCCCGGCAACAATTA	4014
DB	1869	CAACGTTGCGCAAACTATTAACTGCGGAACTACTTACTTCTAGCTTCCCGGCAACAATTA	1928
QY	4015	TAGACTGATGGAGCGGATAAAGTTGCAGAACCACTTCTGCGTTCGCGCTTTCGCGCTG	4074
DB	1929	TAGACTGATGGAGCGGATAAAGTTGCAGAACCACTTCTGCGTTCGCGCTTTCGCGCTG	1988
QY	4075	GCTGGTTTATTCGTGATAAATCTGGAGCGGTGAGCGGTCTCGCGGTATCATTTGAG	4134
DB	1989	GCTGGTTTATTCGTGATAAATCTGGAGCGGTGAGCGGTCTCGCGGTATCATTTGAG	2048
QY	4135	CACCTGGGCGCAGATGGTAAAGCCCTCCCGTATCGTAGTTTATACACGACGGGAGTCAAG	4194
DB	2049	CACCTGGGCGCAGATGGTAAAGCCCTCCCGTATCGTAGTTTATACACGACGGGAGTCAAG	2108
QY	4195	CAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCTGATTAAGCATT	4254
DB	2109	CAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCTGATTAAGCATT	2168
QY	4255	GGTAACTGTCAGACCAAGTTTACTCATATATCTTTAGATTGATTTAAACCTTCAATTTT	4314
DB	2169	GGTAACTGTCAGACCAAGTTTACTCATATATCTTTAGATTGATTTAAACCTTCAATTTT	2228
QY	4315	AATTTAAAGGATCTAGGTGAAGATCTCTTTTGTGATAATCTCATGACCAAAATCCCTTAAC	4374

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 18:49:20 ; Search time 3194 Seconds
(without alignments)
12471.490 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttcacagcttatcatcga.....agttagcgcgaattgatctg 6729

Scoring table: OLIGO NUC

Gapop_60.0 , Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3855	57.3	5027	AAx78865	Aax78865 Human tis
2	3855	57.3	5069	AAx78871	Aax78871 Human tis
3	3855	57.3	5903	ABz23939	Abz23939 Nucleotid
4	3804	56.5	4816	AAx7799	Aax7799 pTRCHisB
5	3804	56.5	4816	ADP72546	Adf72546 Mouse Adi
6	3804	56.5	4816	ADG14708	Adg14708 Globular
7	3804	56.5	5013	ACA63355	Aca63355 E. coli D
8	3804	56.5	5013	AAx59423	Aad59423 E. coli pu
9	3804	56.5	5013	ADE86127	Ade86127 E. coli p
10	3804	56.5	5104	AAx7798	Aax7798 pTRCHisB
11	3804	56.5	5104	ADP72545	Adf72545 Mouse Adi
12	3753	55.8	5069	AAx78874	Aax78874 Human tis
13	3753	55.8	5099	AAx78896	Aax78896 Human tis
14	3753	55.8	5132	AAx78883	Aax78883 Human tis
15	3753	55.8	5324	AAx78890	Aax78890 Human tis
16	3753	55.8	5342	AAx78893	Aax78893 Human tis
17	3753	55.8	5926	AAx32978	Aax32978 Trn7 donor
18	3753	55.8	5926	AAx32977	Aax32977 Trn7 donor
19	3753	55.8	5926	AAx45060	Aad45060 pBM delta
20	3753	55.8	5926	AAx45059	Aad45059 Transposo

C	21	3753	55.8	5926	12	ADG46817	Adg46817 Donor pla
	22	3703	55.0	4214	3	AAA59260	Aaa59260 Nucleotid
	23	3703	55.0	4214	3	AAA62632	Aaa62632 Vector pB
	24	3702	55.0	4176	1	AAx90709	Aan90709 Sequence
	25	3702	55.0	4176	12	ADI36476	Adi36476 Plasmid C
	26	3702	55.0	4203	6	ABQ74925	Abq74925 Avidity I
	27	3702	55.0	4204	6	ABQ74926	Abq74926 Avidity I
	28	3702	55.0	4205	6	ABQ74927	Abq74927 Avidity I
	29	3702	55.0	4216	6	ABQ74928	Abq74928 Avidity I
	30	3702	55.0	4217	6	ABQ74929	Abq74929 Avidity I
	31	3702	55.0	4218	6	ABQ74930	Abq74930 Avidity I
C	32	3702	55.0	5926	12	ADG46818	Adg46818 pBM DNA (
	33	3683	54.7	5103	12	ADG14707	Adg14707 Mouse OB
	34	2741	40.7	5772	10	AAx55822	Aal55822 Phi-29 po
	35	2741	40.7	5772	10	AAx55823	Aal55823 Phi-29 ex
	36	2731	40.6	4145	12	ADO37103	Ado37103 Vector pB
	37	2731	40.6	4145	12	ADO27625	Ado27625 pBAD/gIII
	38	2729	40.6	4357	4	AAF31389	Aaf31389 Expressio
	39	2679	39.8	4189	2	AAQ05397	Aaq05397 Secretion
	40	2661	39.5	6886	6	AAx48883	Aal48883 B subtili
	41	2661	39.5	6886	6	AAx41227	Aal41227 Pantothen
	42	2658	39.5	5797	12	ADL92656	Adl92656 Expressio
	43	2658	39.5	5797	13	ADR42002	Adr42002 VAP relat
	44	2648	39.4	5021	3	AAA07775	Aaa07775 DNA seque
	45	2648	39.4	7268	8	ABz76630	Abz76630 HIF-1 alp

ALIGNMENTS

RESULT 1
AAx78865
ID AAx78865 standard; DNA; 5027 BP.
XX
AC AAx78865;
XX
DT 08-SEP-1999 (first entry)
XX
DE Human tissue factor clone Nuv120 DNA.
XX
KW Tissue factor; human; thrombogenic; substructure; thrombose; tumour;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 413..1171
FT /product= a
FT /tag= a
FT /product= "TF clone Nuv120"
XX
PN WO9932143-Al.
XX
PD 01-JUL-1999.
XX
PF 22-DEC-1998; 98WO-US027498.
XX
PR 23-DEC-1997; 97US-00996744.
XX
PA (NUVA-) NUVAS LLC.
XX
PI Houston LL, Dickinson CD;
XX
PI WPI: 1999-405116/34.
XX
PI P-PSDB; AAY25403.
XX
PT New thrombogenic polypeptides used to, e.g. obliterate vasculative
XX
PS malformations.
XX
PS Claim 30; Page 68-70; 97pp; English.
XX
CC This invention describes novel thrombogenic polypeptides which comprise a
XX
CC thrombogenic substructure and a context-dependent entity which recognizes
XX
CC desired biologically susceptible sites, e.g. tumour vascular endothelium.

CC A novel context-dependent functional entity comprises a substructure with
 CC thrombogenic potential and one or more context-enhancing substructures
 CC having the ability to recognize desired biologically susceptible sites,
 CC where the entity imparts thrombogenic activity when positioned in the
 CC function-forming-context at the biologically susceptible sites, and the
 CC entity has no thrombogenic activity absent a function-forming-context at
 CC the biologically susceptible sites. The context-dependent functional
 CC entities impart thrombogenic activity only at biologically susceptible
 CC sites. They can be used to obliterate vasculature malformations or to
 CC selectively thrombose the vasculature of solid tumours. This sequence
 CC encodes the human tissue factor protein Nu120 which is used in the
 XX method of the invention

SQ Sequence 5027 BP; 1287 A; 1221 C; 1321 G; 1198 T; 0 U; 0 Other;

Query Match		57.3%;	Score 3855;	DB 2;	Length 5027;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 3855;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	2875	AGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTTAAATCAGAAC	2934		
Db	1173	AGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTTAAATCAGAAC	1232		
QY	2935	GCAGAAGCGGTCTGATAAACAAGATTTGCCTGGCGGAGTAGCGCGTGGTCCACCTG	2994		
Db	1233	GCAGAAGCGGTCTGATAAACAAGATTTGCCTGGCGGAGTAGCGCGTGGTCCACCTG	1292		
QY	2995	ACCCCATGCCAACTCAGAAAGTGAACCGCGTAGCGCGATGGTAGTGTGGGCTCTCCCC	3054		
Db	1293	ACCCCATGCCAACTCAGAAAGTGAACCGCGTAGCGCGATGGTAGTGTGGGCTCTCCCC	1352		
QY	3055	ATGCAGAGTAGGGAACCTGCAGGCATCAAAATAAAGAAAGGCTCAGTCGAAAGACTGG	3114		
Db	1353	ATGCAGAGTAGGGAACCTGCAGGCATCAAAATAAAGAAAGGCTCAGTCGAAAGACTGG	1412		
QY	3115	GCCTTTCTGTTTATCTGTTGTTTGTGTCGTTGACCGCTCTCCTGAGTAGGACAAATCCGCG	3174		
Db	1413	GCCTTTCTGTTTATCTGTTGTTTGTGTCGTTGACCGCTCTCCTGAGTAGGACAAATCCGCG	1472		
QY	3175	GGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGTTGGCGGCGAGGACGCCCGCA	3234		
Db	1473	GGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGTTGGCGGCGAGGACGCCCGCA	1532		
QY	3235	TAACTGCCAGGCATCAAAATTAAGCAGAGGCGCATCTGACGGATGGCTTTTTCGGTTT	3294		
Db	1533	TAACTGCCAGGCATCAAAATTAAGCAGAGGCGCATCTGACGGATGGCTTTTTCGGTTT	1592		
QY	3295	CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTTCAAAATATGATCCGCTCATGAGAC	3354		
Db	1593	CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTTCAAAATATGATCCGCTCATGAGAC	1652		
QY	3355	AATAACCTGTATTAATGCTTCAATAATATTTGAAAAGGAAGATGATGATTTCAACATT	3414		
Db	1653	AATAACCTGTATTAATGCTTCAATAATATTTGAAAAGGAAGATGATGATTTCAACATT	1712		
QY	3415	TCCGTGTCGCCCTTATTCCTTTTGTGGGCAATTTTGGCTTCTGTTTGTCTCACCCAG	3474		
Db	1713	TCCGTGTCGCCCTTATTCCTTTTGTGGGCAATTTTGGCTTCTGTTTGTCTCACCCAG	1772		
QY	3475	AAACGCTGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACAGTGGGTATACATCG	3534		
Db	1773	AAACGCTGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACAGTGGGTATACATCG	1832		
QY	3535	AACTGGATCTCAACAGCGGTAAAGTCTTTGAGAGTTTTGCGCCCGGAAAGACGTTTTCCAA	3594		
Db	1833	AACTGGATCTCAACAGCGGTAAAGTCTTTGAGAGTTTTGCGCCCGGAAAGACGTTTTCCAA	1892		
QY	3595	TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACCGCGGC	3654		
Db	1893	TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACCGCGGC	1952		
QY	3655	AAGAGCAACTCGGTGCGCCGCATACACTATTCTCAGAAATGACCTGGTTGAGTACTCACCG	3714		

Db	1953	AAGACCAACTCGGTGCGCGCATACACTATTCTCAGAAATGACTTGGTTGAGTACTCACCG	2012		
QY	3715	TCACAGAAAAAGCATCTTTACGGATGGCATGACAGTAAGAGAATTTATGCAAGTGTGCTCA	3774		
Db	2013	TCACAGAAAAAGCATCTTTACGGATGGCATGACAGTAAGAGAATTTATGCAAGTGTGCTCA	2072		
QY	3775	CCATGAGTGATAACACTGCGGCGCAACTTACTTCTTGACAAAGATCGGAGGACCGAGGAGC	3834		
Db	2073	CCATGAGTGATAACACTGCGGCGCAACTTACTTCTTGACAAAGATCGGAGGACCGAGGAGC	2132		
QY	3835	TAAACCGCTTTTTCACAAATGCGGGATCATGTAACCTGCGCTTGATCGTTGGGAACCGG	3894		
Db	2133	TAAACCGCTTTTTCGACAAATGCGGGATCATGTAACCTGCGCTTGATCGTTGGGAACCGG	2192		
QY	3895	AGCTGAATGAAGCCATACCAACGACGAGCGTGACACCAACGATGCTGTAGCAATGGCAA	3954		
Db	2193	AGCTGAATGAAGCCATACCAACGACGAGCGTGACACCAACGATGCTGTAGCAATGGCAA	2252		
QY	3955	CAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTA	4014		
Db	2253	CAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTA	2312		
QY	4015	TAGACTGATGAGCGCGGATAAAGTTGACGACCACTTCTGCGCTCGGCGCTTCCGGCTG	4074		
Db	2313	TAGACTGATGAGCGCGGATAAAGTTGACGACCACTTCTGCGCTCGGCGCTTCCGGCTG	2372		
QY	4075	GCTGGTTTATTTGCTGATAAAATCTGAGCGCGTGAGCGGTCTGCGCGGTATCATTTGAG	4134		
Db	2373	GCTGGTTTATTTGCTGATAAAATCTGAGCGCGTGAGCGGTCTGCGCGGTATCATTTGAG	2432		
QY	4135	CACCTGGGCGCAGATGTTAAGCCCTCCGCTATGCTAGTTATCTACACGACGGGAGTCAAG	4194		
Db	2433	CACCTGGGCGCAGATGTTAAGCCCTCCGCTATGCTAGTTATCTACACGACGGGAGTCAAG	2492		
QY	4195	CAACTATGGAATGAACAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCAT	4254		
Db	2493	CAACTATGGAATGAACAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCAT	2552		
QY	4255	GGTAACTGTCAGACCAAGTTTACTCATATATATCTTTAGATTGATTTTAAATCTTCAATTT	4314		
Db	2553	GGTAACTGTCAGACCAAGTTTACTCATATATATCTTTAGATTGATTTTAAATCTTCAATTT	2612		
QY	4315	AATTTAAAGGATCTAGGTGAAGATCCCTTTTGTGATAATCTCATGACCAAAATCCCTTAAC	4374		
Db	2613	AATTTAAAGGATCTAGGTGAAGATCCCTTTTGTGATAATCTCATGACCAAAATCCCTTAAC	2672		
QY	4375	GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTTCTTGAG	4434		
Db	2673	GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTTCTTGAG	2732		
QY	4435	ATCCTTTTTCGTTGCGCGTAAATCTGCTGCTTGCAAAACAAAACCAACCGCTACCAGCGG	4494		
Db	2733	ATCCTTTTTCGTTGCGCGTAAATCTGCTGCTTGCAAAACAAAACCAACCGCTACCAGCGG	2792		
QY	4495	TGGTTGTTTTCGCGGATCAAGAGGTACCAACTCTTTTTCGGAAGTAACTGGCTTCAGCA	4554		
Db	2793	TGGTTGTTTTCGCGGATCAAGAGGTACCAACTCTTTTTCGGAAGTAACTGGCTTCAGCA	2852		
QY	4555	GAGCGCAGATACCAAAATCTGCTTCTAGTGTAGCCGCTAGTTAGGCCACCACTTCAAGA	4614		
Db	2853	GAGCGCAGATACCAAAATCTGCTTCTAGTGTAGCCGCTAGTTAGGCCACCACTTCAAGA	2912		
QY	4615	ACTCTGTAGCACCGGCTACATACCTCGCTCTGCTTAATCCTGTACAGTGGCTGTGCCA	4674		
Db	2913	ACTCTGTAGCACCGGCTACATACCTCGCTCTGCTTAATCCTGTACAGTGGCTGTGCCA	2972		
QY	4675	GTGCGGATAAGTGTGTTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC	4734		
Db	2973	GTGCGGATAAGTGTGTTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC	3032		
QY	4735	AGCGGTGCGGCTGAACCGGGGTTGTCGACACAGCCAGCTTGAGCGGAACGACTACA	4794		
Db	3033	AGCGGTGCGGCTGAACCGGGGTTGTCGACACAGCCAGCTTGAGCGGAACGACTACA	3092		

1	3804	56.5	5013	4	US-09-183-1888-5	Sequence 5, Appli
2	3753	55.8	5926	3	US-09-027-169-3	Sequence 3, Appli
3	3753	55.8	5926	3	US-09-027-169-4	Sequence 4, Appli
4	3752	55.8	4476	3	US-08-801-344-2	Sequence 2, Appli
5	3752	55.8	4476	3	US-09-498-599-2	Sequence 2, Appli
6	3729	40.6	4357	4	US-09-795-872-10	Sequence 10, Appli
7	2527	37.6	4593	3	US-08-801-344-1	Sequence 1, Appli
8	2527	37.6	4593	3	US-09-498-599-1	Sequence 1, Appli
9	2433	36.2	5238	6	5453363-1	Patent No. 5453363
10	2433	36.2	5238	6	5453363-1	Patent No. 5453363
11	2138	34.7	6312	1	US-08-531-601-3	Sequence 3, Appli
12	2338	34.7	6312	2	US-08-859-032-3	Sequence 3, Appli
13	2329	34.6	5946	4	US-09-358-856C-1	Sequence 1, Appli
14	2179	32.4	9144	3	US-08-556-9708-79	Sequence 79, Appli
15	2130	31.7	7731	3	US-09-301-593-29	Sequence 29, Appli
16	2130	31.7	7731	3	US-09-301-593-42	Sequence 42, Appli
17	2130	31.7	7892	2	US-07-516-098A-40	Sequence 40, Appli
18	2129	31.6	3699	1	US-08-053-131-120	Sequence 120, App
19	2129	31.6	3699	1	US-08-645-641-120	Sequence 120, App
20	2129	31.6	3699	1	US-07-853-4088-120	Sequence 120, App
21	2129	31.6	3699	1	US-08-096-762-120	Sequence 120, App
22	2129	31.6	3699	2	US-08-308-865-120	Sequence 120, App
23	2129	31.6	3699	5	PCR-US92-10983-120	Sequence 120, App
24	2129	31.6	5115	3	US-08-825-853-19	Sequence 19, Appli
25	2129	31.6	5115	3	US-09-052-888-19	Sequence 19, Appli
26	2129	31.6	5115	4	US-09-723-890-19	Sequence 19, Appli
27	2129	31.6	5115	4	US-09-723-890-19	Sequence 19, Appli

Qy	3175	GGAGCGGANTTTGAACGTTTGGCGAAGCAACGGCCCGGAGGGTGGCGGCGAGACGCCCGGCCA	3234
Db	1459	GGAGCGGANTTTGAACGTTTGGCGAAGCAACGGCCCGGAGGGTGGCGGCGAGACGCCCGGCCA	1518
Qy	3235	TAAACTGCCAGGCATCAAAATTAAAGCAAGAGCCCATCTCTGACGGATGGCCTTTTTCGGTTT	3294
Db	1519	TAAACTGCCAGGCATCAAAATTAAAGCAAGAGCCCATCTCTGACGGATGGCCTTTTTCGGTTT	1578
Qy	3295	CTACAAACTCTTTTTTGTATTATTTTCTTAAATACATTCAAATATGTATCCGCTCATGAGAC	3354
Db	1579	CTACAAACTCTTTTTTGTATTATTTTCTTAAATACATTCAAATATGTATCCGCTCATGAGAC	1638
Qy	3355	AATTAACCTGATAAATCGTTCAATAATTTTGAAGAGGAGATGATGAGTATTCAACATT	3414
Db	1639	AATTAACCTGATAAATCGTTCAATAATTTTGAAGAGGAGATGATGAGTATTCAACATT	1698
Qy	3415	TCCGTGTGCGCCCTTAATCCCTTTTTTTCGGCGCATTTTGCCTTCCTGTTTTTGTCAACCAG	3474
Db	1699	TCCGTGTGCGCCCTTAATCCCTTTTTTTCGGCGCATTTTGCCTTCCTGTTTTTGTCAACCAG	1758
Qy	3475	AAACGCTGGTGAAGATGAAGATGCTGAAGATCAGTTGGGTGCAAGAGTGGGTTCATCG	3534
Db	1759	AAACGCTGGTGAAGATGAAGATGCTGAAGATCAGTTGGGTGCAAGAGTGGGTTCATCG	1818
Qy	3535	AACTGGATCTCAACAGCGGTAAAGATCCTTCAGAGTTTTTCGCCCGGAAGACGTTTTCCAA	3594
Db	1819	AACTGGATCTCAACAGCGGTAAAGATCCTTCAGAGTTTTTCGCCCGGAAGACGTTTTCCAA	1878
Qy	3595	TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGGTPATTATCCGTTGTGACCGCGGC	3654
Db	1879	TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGGTPATTATCCGTTGTGACCGCGGC	1938
Qy	3655	AAGGAGCAACTCGGTGCGGCATACACTATTTCTCAGATGACTTGGTTGAGTACTCACCG	3714
Db	1939	AAGGAGCAACTCGGTGCGGCATACACTATTTCTCAGATGACTTGGTTGAGTACTCACCG	1998
Qy	3715	TCACAGAAAAGCATCTTACGGATGGCATGACGTAAGAGAAATTATGCAGTGTGCCATAA	3774
Db	1999	TCACAGAAAAGCATCTTACGGATGGCATGACGTAAGAGAAATTATGCAGTGTGCCATAA	2058
Qy	3775	CCATGAGTGATAACACTGCGGCCAACTTACTTCTGACACGATCGGAGGACCGAAGAGC	3834
Db	2059	CCATGAGTGATAACACTGCGGCCAACTTACTTCTGACACGATCGGAGGACCGAAGAGC	2118
Qy	3835	TAAACCGCTTTTTTGCACAAATGGGGATCATGTAACTCGCCTTGATCGTTTGGAAACCGG	3894
Db	2119	TAAACCGCTTTTTTGCACAAATGGGGATCATGTAACTCGCCTTGATCGTTTGGAAACCGG	2178
Qy	3895	AGCTGAATGAAGCCATACCAACGACGAGCGTGACACCGATGCCCTGTAGCAATGGCAA	3954
Db	2179	AGCTGAATGAAGCCATACCAACGACGAGCGTGACACCGATGCCCTGTAGCAATGGCAA	2238
Qy	3955	CAACGTTGCGCAAACTAATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAA	4014
Db	2239	CAACGTTGCGCAAACTAATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAA	2298
Qy	4015	TAGACTGGATGAGGCGGATAAAGTTTCGAGGACCACTTCTCGCTCGGCCCTTCCGGCTG	4074
Db	2299	TAGACTGGATGAGGCGGATAAAGTTTCGAGGACCACTTCTCGCTCGGCCCTTCCGGCTG	2358
Qy	4075	GCTGGTTTTATTGCTGATAAATCTGGAGCCGCTGAGCGTGGGTCTCGCGGTATCATTTGCAG	4134
Db	2359	GCTGGTTTTATTGCTGATAAATCTGGAGCCGCTGAGCGTGGGTCTCGCGGTATCATTTGCAG	2418
Qy	4135	CACCTGGGCGCAGATGGTAAGCCCTCCCGTATCTGTAGTTATCTACACGACGGGAGTCAGG	4194
Db	2419	CACCTGGGCGCAGATGGTAAGCCCTCCCGTATCTGTAGTTATCTACACGACGGGAGTCAGG	2478
Qy	4195	CAACTATGGATGAACGAAATAGACAGATTCGCTGAGATGAGTGGCTCTCACTGATTAAAGCAT	4254
Db	2479	CAACTATGGATGAACGAAATAGACAGATTCGCTGAGATGAGTGGCTCTCACTGATTAAAGCAT	2538

QY	4255	GGTTAACTGTCAGACCAAGTTTACTCATATATATCTTTAGATTTGATTTAAACTTCATTTTT	4331
Db	2539	GGTAACTGTCAGACCAAGTTTACTCATATATATCTTTAGATTTGATTTAAAACTTCATTTTT	2598
QY	4315	AAATTAAAGGATCTAGGTGAAGATCCTTTTGTATTAATCTCATGACCAAAATCCCTTAAC	4374
Db	2599	AAATTAAAGGATCTAGGTGAAGATCCTTTTGTATTAATCTCATGACCAAAATCCCTTAAC	2658
QY	4375	GTGAGTTTTCGTTCCACTGAGCGTCAAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG	4434
Db	2659	GTGAGTTTTCGTTCCACTGAGCGTCAAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG	2718
QY	4435	ATCCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACACACGCTACCAAGCGG	4494
Db	2719	ATCCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACACGCTACCAAGCGG	2778
QY	4495	TGTTTTGTTTCCCGGATCAAGAGCTACCAACTCTCTTTTCCGAAGTAACTGGCTTCAGCA	4554
Db	2779	TGTTTTGTTTCCCGGATCAAGAGCTACCAACTCTCTTTTCCGAAGTAACTGGCTTCAGCA	2838
QY	4555	GAGCGCAGATACCAATACCTGCTCTAGTGATGCGGTAGTTTAGGCCACCACTTCAAGA	4614
Db	2839	GAGCGCAGATACCAATACCTGCTCTAGTGATGCGGTAGTTTAGGCCACCACTTCAAGA	2898
QY	4615	ACTCTGTAGCACCGCCCTACATACCTCGCTCTGCTAATCTCTTTTCCAGTGGCTGCTGCCA	4674
Db	2899	ACTCTGTAGCACCGCCCTACATACCTCGCTCTGCTAATCTCTTTTCCAGTGGCTGCTGCCA	2958
QY	4675	GTGCGGATAAGTGTGTCCTTACCGGGTTGACATCAAGACGATAGTTTACCGGATAAGCGCG	4734
Db	2959	GTGCGGATAAGTGTGTCCTTACCGGGTTGACATCAAGACGATAGTTTACCGGATAAGCGCG	3018
QY	4735	AGCGGTCGGCTGAAACGGGGGTTCTGTGCACACAGCCACGCTTTGGAGCGAAACGACCTACA	4794
Db	3019	AGCGGTCGGCTGAAACGGGGGTTCTGTGCACACAGCCACGCTTTGGAGCGAAACGACCTACA	3078
QY	4795	CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAAGGAGAA	4854
Db	3079	CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAAGGAGAA	3138
QY	4855	AGCGGACAGGTATCCGGTAAGCGGAGGTCGGAAACAGGAGAGCCACAGAGGAGCTTC	4914
Db	3139	AGCGGACAGGTATCCGGTAAGCGGAGGTCGGAAACAGGAGAGCCACAGAGGAGCTTC	3198
QY	4915	CAGGGGAAAACGGCTGGTATCTTTTATAGTCTCTGTCGGGTTTCGCCACCTCTGACTTGAGC	4974
Db	3199	CAGGGGAAAACGGCTGGTATCTTTTATAGTCTCTGTCGGGTTTCGCCACCTCTGACTTGAGC	3258
QY	4975	GTCAATTTTGTGATGCTGTCAAGGGGGCGGACCTATGGAATAACGCCAGCAACGCCG	5034
Db	3259	GTCAATTTTGTGATGCTGTCAAGGGGGCGGACCTATGGAATAACGCCAGCAACGCCG	3318
QY	5035	CCTTTTTCACGGTTCCTGGCTTTTGTCTGGCTTTTGTCTCACATGTTCTTTTCCCTGCGTTAT	5094
Db	3319	CCTTTTTCACGGTTCCTGGCTTTTGTCTGGCTTTTGTCTCACATGTTCTTTTCCCTGCGTTAT	3378
QY	5095	CCCCTGATCTGTGGATAAACCGTATTACCGCTTTTGTAGTGAGCTGTATACCGCTCGCCGCA	5154
Db	3379	CCCCTGATCTGTGGATAAACCGTATTACCGCTTTTGTAGTGAGCTGTATACCGCTCGCCGCA	3438
QY	5155	GCCGAAACGACCGAGCGCAGCGAGTCACTGAGCGAGGAAGCGGAGAGCGCTCATGCGGT	5214
Db	3439	GCCGAAACGACCGAGCGCAGCGAGTCACTGAGCGAGGAAGCGGAGAGCGCTCATGCGGT	3498
QY	5215	ATTTTCTCCTTACGCATCTGTGGGTATTTTCAACCGCATATGGTGACCTCTCAGTACAA	5274
Db	3499	ATTTTCTCCTTACGCATCTGTGGGTATTTTCAACCGCATATGGTGACCTCTCAGTACAA	3558
QY	5275	TCTGCTCTGATCGCGATAGTTAAGCCAGTATACCTCCGCTATCGCTACGTGACCTGGGT	5334
Db	3559	TCTGCTCTGATCGCGCATAGTTAAGCCAGTATACCTCCGCTATCGCTACGTGACCTGGGT	3618
QY	5335	CATGGCTCGCCCCGCAACCCGCGCAACACCGCTGTGAGCGCCCTCGACGGGCTTGTCTGCT	5394

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 18:42:20 ; Search time 3776 Seconds
(without alignments)
11551.793 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttcacagcttatcatcga.....agttacgcgaattgatctg 6729

Scoring table: OLIGO NUC

Gapop_60.0 , Gapext 60.0

Searched: 7297361 seqs, 3241162794 residues

Word size : 0

Total number of hits satisfying chosen parameters: 14594722

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6729	100.0	6729	21	US-10-649-433C-1
2	3804	56.5	5013	14	US-10-198-034-5
3	3804	56.5	5013	15	US-10-198-228-5
C 4	3753	55.8	5926	13	US-10-024-809-3
C 5	3753	55.8	5926	13	US-10-024-809-4
6	3702	55.0	4176	16	US-10-244-142A-3
7	2741	40.7	5772	20	US-10-131-998A-37
Sequence 1, Appli					
Sequence 5, Appli					
Sequence 5, Appli					
Sequence 3, Appli					
Sequence 4, Appli					
Sequence 3, Appli					
Sequence 37, Appli					

8	2741	40.7	5772	20	US-10-131-998A-38	Sequence 38, Appli
9	2731	40.6	4145	20	US-10-699-088-163	Sequence 163, App
10	2731	40.6	4145	21	US-10-699-114-163	Sequence 163, App
11	2729	40.6	4357	10	US-09-795-872-10	Sequence 10, Appl
12	2729	40.6	4357	19	US-10-662-824-10	Sequence 10, Appl
13	2661	39.5	6886	18	US-10-466-642-20	Sequence 20, Appl
14	2661	39.5	6886	18	US-10-466-717-20	Sequence 20, Appl
15	2658	39.5	5797	17	US-10-316-194-167	Sequence 167, App
16	2658	39.5	5797	21	US-10-732-180-203	Sequence 203, App
17	2613	38.8	6806	10	US-09-908-943A-194	Sequence 194, App
18	2613	38.8	6806	20	US-10-801-487-194	Sequence 194, App
19	2613	38.8	6806	20	US-10-801-938-194	Sequence 194, App
20	2613	38.8	6806	20	US-10-801-509-194	Sequence 194, App
21	2613	38.8	6806	20	US-10-801-486-194	Sequence 194, App
22	2613	38.8	6806	21	US-10-801-493-194	Sequence 194, App
23	2491	37.0	4921	21	US-10-495-491-2	Sequence 2, Appli
24	2491	37.0	4935	21	US-10-495-491-1	Sequence 1, Appli
25	2491	37.0	4945	21	US-10-495-491-3	Sequence 3, Appli
26	2491	37.0	4951	21	US-10-495-491-4	Sequence 4, Appli
27	2484	36.9	5398	22	US-10-505-082-3	Sequence 3, Appli
28	2329	34.6	5446	17	US-10-393-269-1	Sequence 1, Appli
29	2279	33.9	6464	15	US-10-151-690-20	Sequence 20, Appli
C 30	2260	33.6	7615	10	US-09-898-616A-9	Sequence 9, Appli
C 31	2260	33.6	7615	17	US-10-187-498A-9	Sequence 9, Appli
32	2227	33.1	5522	10	US-09-872-868-16	Sequence 16, Appli
33	2227	33.1	5525	10	US-09-872-868-17	Sequence 17, Appli
34	2227	33.1	6086	10	US-09-872-868-12	Sequence 12, Appli
35	2227	33.1	6086	10	US-09-872-868-13	Sequence 13, Appli
36	2227	33.1	6088	10	US-09-872-868-14	Sequence 14, Appli
37	2227	33.1	6088	10	US-09-872-868-15	Sequence 15, Appli
38	2130	31.7	7731	9	US-09-301-593-29	Sequence 29, Appli
39	2130	31.7	7731	9	US-09-301-593-42	Sequence 42, Appli
40	2130	31.7	7731	15	US-10-159-006-29	Sequence 29, Appli
41	2130	31.7	7731	15	US-10-159-006-42	Sequence 42, Appli
C 42	2130	31.7	9077	9	US-09-734-300-1	Sequence 1, Appli
C 43	2130	31.7	9077	9	US-09-734-300-3	Sequence 1, Appli
C 44	2129	31.6	4010	16	US-10-353-908-1	Sequence 1, Appli
C 45	2129	31.6	5024	17	US-10-313-963A-52	Sequence 52, Appli

ALIGNMENTS

RESULT 1
US-10-649-433C-1
; Sequence 1, Application US/10649433C
; Publication No. US20050118726A1
; GENERAL INFORMATION:
; APPLICANT: Schultz, Jerome S.
; TITLE OF INVENTION: System and Method for Detecting Bioanalytes and Method for
; TITLE OF INVENTION: Producing a Bioanalyte Sensor
; FILE REFERENCE: 03-016
; CURRENT APPLICATION NUMBER: US/10/649,433C
; CURRENT FILING DATE: 2003-08-26
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 6729
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-649-433C-1

Query Match 100.0%; Score 6729; DB 21; Length 6729;					
Best Local Similarity 100.0%; Pred. No. 0;					
Matches 6729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1	GTTCACAGCTTATCATCGACGTGCGGTCACCAATGCTTCTGCGTCAGCAGCATC	60		
Db	1	GTTCACAGCTTATCATCGACGTGCGGTCACCAATGCTTCTGCGTCAGCAGCATC	60		
Qy	61	GGAAGCTGTGGTATGGCTGTGTCAGGTGCGTAAATCACTGCATAATTCGTCGCTCAAGGC	120		
Db	61	GGAAGCTGTGGTATGGCTGTGTCAGGTGCGTAAATCACTGCATAATTCGTCGCTCAAGGC	120		

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 19:32:11 ; Search time 201.5 Seconds
(without alignments)

6426.229 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttgacagcttatcatcg.....agttagcgcgaattgatctg 6729

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565918

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10649433/runat_02082005_165026_1831/app_query.fasta_1.6919
-DB=PIR -QFMT=fastan -SUFFIX=oligo.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CGN 1 1 332 @runat_02082005_165026_1831 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309	14.2	332	1 JG8CG	D-galactose-bindin
2	309	14.2	332	2 B91009	galactose-binding
3	309	14.2	332	2 B85853	galactose-binding
4	284	13.0	360	1 RPECL	lactose operon rep
5	284	13.0	360	4 S58116	lac repressor - sy
6	204	9.4	286	2 T51301	beta-lactamase (EC
7	204	9.4	286	4 S41975	beta-lactamase (EC
8	204	9.4	286	1 PNECP	beta-lactamase (EC
9	185	8.5	286	1 I40905	beta-lactamase (EC
10	185	8.5	286	4 I40905	beta-lactamase (EC
11	183	8.4	360	2 B85529	lac operon transcr
12	183	8.4	360	2 F90678	lac operon transcr
13	104	4.8	105	2 JC2566	bla protein - Pseu
14	104	4.8	286	2 S60312	extended spectrum

C	15	90	4.1	149	2	JQ1541	hypothetical 16.9K
	16	86	3.9	286	2	JQ1546	Bla protein - Salm
	17	86	3.9	332	1	S29390	galactose-binding
	18	86	3.9	332	2	AF0781	D-galactose-bindin
C	19	83	3.8	84	1	Q08C7	hypothetical prote
	20	79	3.6	286	2	S30113	beta-lactamase (EC
	21	78	3.6	238	1	JQ1514	green-fluorescent
C	22	70	3.2	91	2	A27354	hypothetical prote
	23	64	2.9	286	2	S60310	extended spectrum
	24	64	2.9	286	2	S60311	beta-lactamase (EC
C	25	62	2.8	88	1	Q08CC8	hypothetical prote
	26	60	2.8	332	1	S15554	D-galactose-bindin
	27	53	2.4	335	2	AG0183	galactose-binding
	28	36	1.7	42	1	Q08C45	hypothetical prote
C	29	32	1.5	293	2	E64874	probable metal-dep
C	30	32	1.5	293	2	F90858	probable enzymes
C	31	32	1.5	293	2	A85761	probable enzymes y
	32	28	1.3	313	2	A48903	beta-lactamase - p
C	33	22	1.0	293	2	AG0653	conserved hypothet
	34	18	0.8	1630	2	A53577	ascites stialoglyco
C	35	17	0.8	54	2	S12504	mb-1 protein - mou
	36	17	0.8	324	2	C82214	galactoside ABC tr
	37	16	0.7	349	1	G64096	D-galactose-bindin
	38	15	0.7	306	1	B45822	beta-lactamase (EC
C	39	15	0.7	507	2	I40767	catalase (EC 1.11.
	40	14	0.6	231	2	JC7702	NZ-3 antigen - hum
C	41	14	0.6	493	2	A55092	catalase (EC 1.11.
	42	13	0.6	298	2	A41381	beta-lactamase (EC
C	43	13	0.6	539	2	S00067	xylan 1,4-beta-xyl
	44	12	0.5	63	1	RGECRE	regulatory protein
C	45	12	0.6	279	2	A24469	beta-lactamase (EC

ALIGNMENTS

RESULT 1

JG8CG

D-galactose-binding periplasmic protein precursor - Escherichia coli (strain K-12)
N:Alternate names: mglB protein
C:Species: Escherichia coli
C>Date: 02-Apr-1982 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A37277; A92319; A94599; A30384; A32653; E64983; A03427; Q00294
R:Hogg, R.W.; Voelker, C.; Von Carlowitz, I.
Mol. Gen. Genet. 229, 453-459, 1991
A>Title: Nucleotide sequence and analysis of the mgl operon of Escherichia coli K12.
A:Reference number: A37277; MUID:92049246; PMID:1719366
A:Accession: A37277
A:Molecule type: DNA
A:Residues: 1-332 <HOG>
A:Cross-references: UNIPROT:P02927; GB:M59444; NID:g146852; PIDN:AAA24169.1; PID:g14685.
A:Experimental source: strain K12
R:Mahoney, W.C.; Hogg, R.W.; Hermanson, M.A.
J. Biol. Chem. 256, 4350-4356, 1981
A>Title: The amino acid sequence of the D-galactose-binding protein from Escherichia coli
A:Reference number: A32319; MUID:81168234; PMID:7012152
A:Accession: A92319
A:Molecule type: protein
A:Residues: 24-178, 'KE', 181-332 <MA1>
A:Experimental source: strain B/r
A>Note: this sequence revised in reference A94599
R:Mahoney, W.C.; Hogg, R.W.; Hermanson, M.A.
submitted to the Atlas, November 1982
A:Reference number: A94599
A:Contents: revisions
A:Accession: A94599
A:Molecule type: protein
A:Residues: 24-332 <MA2>
R:Scholle, A.; Vreemann, J.; Blank, V.; Nold, A.; Boos, W.; Manson, M.D.
Mol. Gen. Genet. 208, 247-253, 1987
A>Title: Sequence of the mglB gene from Escherichia coli K12: comparison of wild-type a

A:Reference number: A30384; MUID:87286407; PMID:3302609

A:Accession: A30384

A:Molecule type: DNA

A:Residues: 1-15 <SCH>
R:Scripture, J.B.; Hogg, R.W.
J. Biol. Chem. 258, 10853-10855, 1983
A:Title: The nucleotide sequences defining the signal peptides of the galactose-binding
A:Reference number: A32653; MUID:83291030; PMID:6885805
A:Accession: A32653
A:Molecule type: DNA
A:Residues: 1-15 <SCR>
R:Vyas, N.K.; Vyas, M.N.; Quijcho, F.A.
Proc. Natl. Acad. Sci. U.S.A. 80, 1792-1796, 1983
A:Title: The 3 angstrom resolution structure of a D-galactose-binding protein for transp
A:Reference number: A3942; MUID:83169767; PMID:6340108
A:Contents: annotation; X-ray crystallography, 3.0 angstroms
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64983
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-332 <BLAT>
A:CROSS-references: GB:AE000304; GB:U00096; NID:gl788470; PIDN:AAC75211.1; PID:gl788473;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: mglB; mglM
A:Map position: 46 min
A:Superfamily: D-galactose-binding protein
C:Keywords: calcium binding; chemotaxis; periplasmic space; sugar transport
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-332/Product: D-galactose-binding protein #status experimental <MAT>
F:37.114,175,177,181,234,259,279/Binding site: galactose (Asp, Asn, His, Asp, Arg, Asn, e
F:157,159,161,163,165,228/Binding site: calcium (Asp, Asn, Asp, Gln, Gln, Glu) #status e

Alignment Scores:
Pred. No.: 3 98e-308 Length: 332
Score: 309.00 Matches: 309
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.17% Indels: 0
DB: 1 Gaps: 0

US-10-649-433C-1 (1-6729) x JGECG (1-332)

QY 1136 GCTGATCTCGCATTCGATGTAACAATCTAATAGTACGACGATACTTATGCTCTAGTG 1195
Db 24 AlaAspThrArgIleGlyValThrIleTyrLysTyrAspAspAsnPheMetSerValVal 43
QY 1196 CCACAGGCTATTGACACAGATCGAAGCCGCGCCAGATGTTTCAGCTGCTGATGATGAT 1255
Db 44 ArgLysAlaIleGluGlnAspAlaLysAlaAlaProAspValGlnLeuLeuMetAsnAsp 63
QY 1256 TCTCAGATGACCACTCCAAAGCAGAACGATCAGATCAGATGCTGCTGCTGCGCAAGGGGTG 1315
Db 64 SerGlnAsnAspGlnSerLysGlnAsnAspGlnIleAspValLeuLeuAlaLysGlyVal 83
QY 1316 AAGCAGCTGGCCATCAACCTGTTGACCCGCGCAGCTGCGGTACGGTGAATGAGAAGCG 1375
Db 84 LysAlaLeuAlaIleAsnLeuValAspProAlaAlaAlaGlyThrValIleGluLysAla 103
QY 1376 CGTGGGCAACGTCGGGTGGTGTCTTCACAAAGAACCGTCTCGTAAGCGCTGGAT 1435
Db 104 ArgGlyGlnAsnValProValValPhePheAsnLysGluProSerArgLysAlaLeuAsp 123
QY 1436 AGCTACGACAAAGCCCTACTACCTTGGCAGTCACTCAAAAGAGTCGCGCATATTTCAGGC 1495
Db 124 SerTyrAspLysAlaTyrTyrValGlyThrAspSerLysGluSerGlyIleIleGlnGly 143
QY 1496 GATTTGATGCTAAACATCTGGCGCGCAATCAGGTGGATCTCGAACAAAGACGGTCAG 1555
Db 144 AspLeuIleAlaLysHisTrpAlaAlaAsnGlnGlyTyrAspLeuAsnLysAspGlyGln 163
QY 1556 ATTCACTGCTACTGCTGAAAGGTGAAACCGGGCCATCCGGATGCGAAGCAGTACCCT 1615

Db 164 IleGlnPheValLeuLeuLysGlyGluProGlyHisProAspAlaGluAlaArgThrThr 183
QY 1616 TACCTGATTAAAGAAATTGAACATGAAGGCATCAAAACTGAACAGTTACAGTTAGATACC 1675
Db 184 TyrValIleLysGluLeuAsnAspLysGlyIleLysThrGluGlnLeuGlnLeuAspThr 203
QY 1676 GCAATGTGGGACACCCCTCAGCGGAAGATGAAGATGAGCGCTGCTCTGCGCCCGAAC 1735
Db 204 AlaMetTrpAspThrAlaGlnAlaLysAspLysMetAspAlaTrpLeuSerGlyProAsn 223
QY 1736 GCCAACAAAATCGAAGTGGTTATCGCAACACCATGCGATGCGAATGGCGCGCTGCAA 1795
Db 224 AlaAsnLysIleGluValValIleAlaAsnAsnAspAlaMetAlaMetGlyAlaValGlu 243
QY 1796 CGCTGAAAGCACACCAACATCCAGCATTCGGGTGTTTGGCGTCGATGCGCTGCAGAA 1855
Db 244 AlaLeuLysAlaHisAsnLysSerSerIleProValPheGlyValAspAlaLeuProGlu 263
QY 1856 CGCTGGCGCTGGTGAATCCGGTGCATCTGCGGCGCACCGTACTGAACGATGCTTAACAAC 1915
Db 264 AlaLeuAlaLeuValLysSerGlyAlaLeuAlaGlyThrValLeuAsnAspAlaAsnAsn 283
QY 1916 CAGCGAAAGCGACCTTTGATCTGCGGAAAACCTGCGCGCATGGTAAAGGTGCGGCTGAT 1975
Db 284 GlnAlaLysAlaThrPheAspLeuAlaLysAsnLeuAlaAspGlyLysGlyAlaAlaAsp 303
QY 1976 GGCACCACTGGAAAAATCGACAAACAGTGTGCGCGTACCTTATGTTGGCGTAGATAA 2035
Db 304 GlyThrAsnTrpLysIleAspAsnLysValValArgValProTyrValGlyValAspLys 323
QY 2036 GACACCTGGCTGAATTCAGCAAGAA 2062
Db 324 AspAsnLeuAlaGluPheSerLysLys 332

RESULT 2
B31009
galactose-binding transport protein EC3042 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: B91009
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:CROSS-references: UNIPROT:Q8X641; GB:BA000007; PIDN:BA036465.1; PID:gl3362511; GSPDB:1-C
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC3042
C:Superfamily: D-galactose-binding protein

Alignment Scores:
Pred. No.: 3 98e-308 Length: 332
Score: 309.00 Matches: 309
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.17% Indels: 0
DB: 2 Gaps: 0

US-10-649-433C-1 (1-6729) x B91009 (1-332)

QY 1136 GCTGATCTCGCATTCGATGTAACAATCTAATAGTACGACGATACTTATGCTCTAGTG 1195
Db 24 AlaAspThrArgIleGlyValThrIleTyrLysTyrAspAspAsnPheMetSerValVal 43
QY 1196 CGCAGGCTATTGACACAGATCGAAGCCGCGCCAGATGTTTCAGCTGCTGATGATGAT 1255
Db 44 ArgLysAlaIleGluGlnAspAlaLysAlaAlaProAspValGlnLeuLeuMetAsnAsp 63
QY 1256 TCTCAGATGACCACTCCAAAGCAGAACGATCAGATCAGATGCTGCTGCTGCGCAAGGGGTG 1315
Db 64 SerGlnAsnAspGlnSerLysGlnAsnAspGlnIleAspValLeuLeuAlaLysGlyVal 83
QY 1316 AAGCAGCTGGCCATCAACCTGTTGACCCGCGCAGCTGCGGTACGGTGAATGAGAAGCG 1375
Db 84 LysAlaLeuAlaIleAsnLeuValAspProAlaAlaAlaGlyThrValIleGluLysAla 103
QY 1376 CGTGGGCAACGTCGGGTGGTGTCTTCACAAAGAACCGTCTCGTAAGCGCTGGAT 1435
Db 104 ArgGlyGlnAsnValProValValPhePheAsnLysGluProSerArgLysAlaLeuAsp 123
QY 1436 AGCTACGACAAAGCCCTACTACCTTGGCAGTCACTCAAAAGAGTCGCGCATATTTCAGGC 1495
Db 124 SerTyrAspLysAlaTyrTyrValGlyThrAspSerLysGluSerGlyIleIleGlnGly 143
QY 1496 GATTTGATGCTAAACATCTGGCGCGCAATCAGGTGGATCTCGAACAAAGACGGTCAG 1555
Db 144 AspLeuIleAlaLysHisTrpAlaAlaAsnGlnGlyTyrAspLeuAsnLysAspGlyGln 163
QY 1556 ATTCACTGCTACTGCTGAAAGGTGAAACCGGGCCATCCGGATGCGAAGCAGTACCCT 1615

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2005, 01:28:36 ; Search time 19524 Seconds
(without alignments)
13118.955 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttgacagcttatcatcg.....agttagcggaattgatctg 6729

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 34239544 seqs, 19032134700 residues

Word size : 0

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	864	12.8	1070	1	AJ281552 4A3A-P6F1
C 2	788	11.7	841	1	AL042026 DKF2p434E
C 3	770	11.4	865	7	CK125894 BES182411
C 4	742	11.0	793	5	BQ751655 EST632218
C 5	742	11.0	806	5	BQ751220 EST631783
C 6	742	11.0	827	7	CN823902 Oa splbn
C 7	741	11.0	758	9	CL422788 AE0544 Sa
C 8	736	10.9	856	7	CN823189 Oa splbn
C 9	732	10.9	928	7	CO487414 GQ0227.B7
C 10	723	10.7	1004	1	AJ281480 4A3A-P4G8
C 11	722	10.7	789	6	CD280920 G44224.42
C 12	717	10.7	769	7	CK118014 218n06.p1
C 13	706	10.5	759	6	CD279661 G43818.35
C 14	699	10.4	780	5	BQ825693 1030129B0
C 15	699	10.4	846	7	CV468077 est_l_van
C 16	690	10.3	741	6	CD279174 G44221.83
C 17	679	10.1	730	6	CD281097 G44224.38
C 18	678	10.1	729	6	CD279546 G44222.81
C 19	678	10.1	800	1	AJ281449 4A3A-P4D5
C 20	677	10.1	728	6	CD279322 G44223.13
C 21	675	10.0	726	6	CD281811 G44471.30
C 22	674	10.0	954	1	AL044364 DKF2p434C
C 23	673	10.0	710	7	CK394569 hggada5A0
C 24	671	10.0	733	8	BZ049372 jnr50b10.

C	25	671	10.0	779	7	CK394523 hggada4D0
	26	666	9.9	727	7	CR791051 DKF2p469B
	27	665	9.9	794	5	BQ751936 EST632499
	28	657	9.8	708	6	CD281616 G44471.18
	29	655	9.7	706	7	CR543261 DKF2p470P
	30	652	9.7	703	6	CD279544 G44222.73
	31	652	9.7	832	7	CN822433 Oa splbn
	32	649	9.6	700	5	BQ660293 H01N05W
	33	648	9.6	699	5	BQ659852 HF01022W
	34	648	9.6	700	6	CA305674 EST000022
	35	648	9.6	819	6	CD849375 CV6nd0008
	36	644	9.6	697	5	BQ660197 H01G17W
	37	644	9.6	697	5	BQ660261 H01L03W
	38	644	9.6	793	7	CK122681 BES182410
	39	641	9.5	641	4	BJ063872 BJ063872
	40	640	9.5	1013	4	BM438846 IPLvR0015
	41	638	9.5	697	7	CK739326 EA0282 Da
	42	638	9.5	832	4	BG23768 602825893
	43	636	9.5	769	7	CV224987 CS hyp_24
	44	636	9.5	786	7	CN823164 Oa splbn
	45	636	9.5	806	4	BJ684280 BJ684280

ALIGNMENTS

RESULT 1	AJ281552/c	AJ281552	1070 bp	mRNA	linear	EST 30-JUN-2000
LOCUS	4A3A-P6F11-F	Anopheles gambiae	immune competent	4A3A	Anopheles	
DEFINITION	gambiae cDNA clone 4A3A-P6F11, mRNA sequence.					
ACCESSION	AJ281552					
VERSION	AJ281552.1	GI:6929432				
KEYWORDS	EST					
SOURCE	Anopheles gambiae (African malaria mosquito)					
ORGANISM	Anopheles gambiae					
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.					
REFERENCE	1 (bases 1 to 1070)					
AUTHORS	Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C., Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B. and Kafatos,F.C.					
TITLE	Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines					
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)					
MEDLINE	20300950					
PUBMED	10841561					
COMMENT	Contact: Dimopoulos G Fotis C. Kafatos laboratory European Molecular Biology Laboratory Meyerhofstrasse 1, 69117 Heidelberg, Germany. Location/Qualifiers 1..1070 /organism="Anopheles gambiae" /mol_type="mRNA" /strain="4A r/r" /db_xref="taxon:7165" /clone="4A3A-P6F11" /cell_line="immune competent 4A3A" /lab_host="E. coli DH10B" /note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI; Site 2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Ronaldo, Lennon & Soares (1996) : Normalization and Subtraction: two approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."					
FEATURES	source					
ORIGIN						

Query Match	12.8%;	Score 864;	DB 1;	Length 1070;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 964;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
4064	CTTCCGGCTGGCTGGTTATTGCTGATAAATCTGGAGCCGGTCAGCGTGGTCTCCGGG	4123		
966	CTTCCGGCTGGCTGGTTATTGCTGATAAATCTGGAGCCGGTCAGCGTGGTCTCCGGG	907		
4124	TATCATTTGACGACCTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGAC	4183		
906	TATCATTTGACGACCTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGAC	847		
4184	GGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGTGCCTCACT	4243		
846	GGGAGTCAGGCAACTATGGATGGAGCAAAATAGACAGATCGCTGAGATAGTGCCTCACT	787		
4244	GATTAAAGCATTTGGTAACTGTGACAGCAAGTTTACTCATATATACATTTAGATTGATTTAAA	4303		
786	GATTAAAGCATTTGGTAACTGTGACAGCAAGTTTACTCATATATACATTTAGATTGATTTAAA	727		
4304	ACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTGATAATCTCATGACCAA	4363		
726	ACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTGATAATCTCATGACCAA	667		
4364	AATCCCTTAACTGAGTTTTCGTTCCACTGAGCGTCAGACCCGCTGAGAAAGATCAAAGG	4423		
666	AATCCCTTAACTGAGTTTTCGTTCCACTGAGCGTCAGACCCGCTGAGAAAGATCAAAGG	607		
4424	ATCTTCTTTGAGATCCTTTTTCGCGGTAATCTGCTGCTTGCAGAAACAAAAACCAACC	4483		
606	ATCTTCTTTGAGATCCTTTTTCGCGGTAATCTGCTGCTTGCAGAAACAAAAACCAACC	547		
4484	GCTACCGCGGTGGTTTGTGGCGGATCAAGAGCTACCAACTCTTTTTCGGAAGTTAAC	4543		
546	GCTACCGCGGTGGTTTGTGGCGGATCAAGAGCTACCAACTCTTTTTCGGAAGTTAAC	487		
4544	TGGCTTCAGCAGAGCGCAGATACCAAAATCTGTCCTTCTAGTGATAGCGGTAGTTAGGCCA	4603		
486	TGGCTTCAGCAGAGCGCAGATACCAAAATCTGTCCTTCTAGTGATAGCGGTAGTTAGGCCA	427		
4604	CCACTTCAAGAACTCTGTAGCAGCCGCTACATACCTCGCTCTGCTAACTCTGTTACCACT	4663		
426	CCACTTCAAGAACTCTGTAGCAGCCGCTACATACCTCGCTCTGCTAACTCTGTTACCACT	367		
4664	GGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTGGACTCAAGACGATAGTTACC	4723		
366	GGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTGGACTCAAGACGATAGTTACC	307		
4724	GGATAAGCGCAGCGTGGGCTGAAACGGGGGGTTTCGTGCACACAGCCCGCTTGGAGCG	4783		
306	GGATAAGCGCAGCGTGGGCTGAAACGGGGGGTTTCGTGCACACAGCCCGCTTGGAGCG	247		
4784	AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAGCTTCC	4843		
246	AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAGCTTCC	187		
4844	CGAAGGAGAAAGCGGACAGGTATCCGGTAAAGCGGCGAGGTCGGAACAGGAGACGCAC	4903		
186	CGAAGGAGAAAGCGGACAGGTATCCGGTAAAGCGGCGAGGTCGGAACAGGAGACGCAC	127		
4904	GAGGAGCTTCCAGGGGAGAACCGCTGGTATCTTTTATAGTCTGTCGGGTTTCGCCACCT	4963		
126	GAGGAGCTTCCAGGGGAGAACCGCTGGTATCTTTTATAGTCTGTCGGGTTTCGCCACCT	67		
4964	CTGACTTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCCTATGAAAAACGC	5023		
66	CTGACTTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCCTATGAAAAACGC	7		
5024	CAGCAA 5029			
6	CAGCAA 1			

RESULT 2	AL042026	841 bp	mRNA	linear	EST 04-SEP-2003
LOCUS	DKFZp434E111.r1.434 (synonym: htes3)	Homo sapiens cDNA clone			
DEFINITION	DKFZp434E111_5', mRNA sequence.				
ACCESSION	AL042026				
VERSION	AL042026.1	GI:5421372			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 841)				
TITLE	Poustka,A., Klein,M., Mewes,H.W., Gassenhuber,J. and Wiemann,S.				
JOURNAL	EST (Poustka, et al.)				
COMMENT	Unpublished (1999)				
	Contact: MIPS				
	MIPS				
	Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany				
	This is the 5' sequence of the clone insert				
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;				
	sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.				
	No sl sequence available.				
	This clone (DKFZp434E111) is available at the RZPD in Berlin.				
	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.				
FEATURES	Location/Qualifiers				
source	1..841				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="DKFZp434E111"				
	/tissue_type="testis"				
	/dev_stage="adult"				
	/lab_host="DH10B"				
	/clone_lib="434 (synonym: htes3)"				
ORIGIN	/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"				
	Query Match 11.7%; Score 788; DB 1; Length 841;				
	Best Local Similarity 99.9%; Pred. No. 0;				
	Matches 838; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	3307	TTTGTGTTATTTTCTAAATACATTCAAATATGTTATCCGCTCATGAGACAATAACCTGAT	3366		
Db	3	TTTGTGTTATTTTCTAAATACATTCAAATATGTTATCCGCTCATGAGACAATAACCTGAT	62		
QY	3367	AAATGCTTCAATAATATTGAAAAAGGAAGATGAGTATGAGTATTCACATTTCCGTGTCGCC	3426		
Db	63	AAATGCTTCAATAATATTGAAAAAGGAAGATGAGTATTCACATTTCCGTGTCGCC	122		
QY	3427	TTATTCCTTTTTCGCGCATTTTTCCTCTCTGTTTTCCTCACCACCAAGACGTTGTA	3486		
Db	123	TTATTCCTTTTTCGCGCATTTTTCCTCTCTGTTTTCCTCACCACCAAGACGTTGTA	182		
QY	3487	AAGTAAAGATGCTGAAGATCAGTTTCGGGTGCACGAGTGGGTTTACATCGAACTCGATCTCA	3546		
Db	183	AAGTAAAGATGCTGAAGATCAGTTTCGGGTGCACGAGTGGGTTTACATCGAACTCGATCTCA	242		
QY	3547	ACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCGGAAGACGTTTTCATATGATGAGCACTT	3606		
Db	243	ACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCGGAAGACGTTTTCATATGATGAGCACTT	302		
QY	3607	TTAAAGTTCTGCTATGTCGGCGGTATTTATCCCGTTGTCGCCGGGCAAGCAACTCG	3666		
Db	303	TTAAAGTTCTGCTATGTCGGCGGTATTTATCCCGTTGTCGCCGGGCAAGCAACTCG	362		
QY	3667	GTCGCCGCATACACTATTCTCAGAAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGC	3726		
Db	363	GTCGCCGCATACACTATTCTCAGAAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGC	422		

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 19:25:21 ; Search time 1028 Seconds

(without alignments)

6703.854 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttagacgttatcatcg.....agtttagcggaattgatctg 6729

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3224408

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV-xlp
-MODEL=frame+ n2p.model -DEV-xlp

-Q/cgn2_1/USPTO_spool_P/US10649433/runat_02082005_165025_1820/app_query.fasta_1.6919

-DB=UniProt -QFWT=fastan -SUFFIX=Oligo.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc

-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10649433 @CGN 1 1 1271 @runat_02082005_165025_1820 -NCPU=6 -ICPU=3

-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	309	14.2	332	1 DGAL_ECOLI	P02927 escherichia
2	309	14.2	332	2 Q8X641	Q8X641 escherichia
3	284	13.0	360	1 LACI_ECOLI	P03023 escherichia
4	222	10.2	286	2 Q91UY8	Q91UY8 plasmid pww
5	204	9.4	286	2 Q38058	Q38058 bacterioph
6	204	9.4	286	2 Q00626	Q00626 staphylococ
7	204	9.4	286	2 Q79CL6	Q79CL6 methylobaci
8	204	9.4	286	2 Q79D93	Q79D93 escherichia
9	203	9.3	285	2 Q6KB67	Q6KB67 hordeum vul
10	185	8.5	286	1 BLAT_ECOLI	P62594 salmonella
11	185	8.5	286	1 BLAT_SALTI	P62594 salmonella
12	185	8.5	286	2 Q6A253	Q6A253 haemophilus
13	185	8.5	286	2 Q6LBN9	Q6LBN9 pseudomonas
14	185	8.5	286	2 Q6LCV6	Q6LCV6 neisseria g
15	185	8.5	286	2 Q6TMH1	Q6TMH1 streptococc
16	185	8.5	286	2 Q6W9J1	Q6W9J1 enterobacte

17	185	8.5	286	2 Q6WRX2	Q6WRX2 zymomonas m
18	185	8.5	286	2 Q6WZD4	Q6WZD4 acinetobact
19	185	8.5	286	2 Q7B3X5	Q7B3X5 citrobacter
20	185	8.5	286	2 Q7B899	Q7B899 klebsiella
21	185	8.5	286	2 Q7BP57	Q7BP57 shigella fl
22	185	8.5	286	2 Q7BR75	Q7BR75 neisseria m
23	185	8.5	286	2 Q7DFY3	Q7DFY3 salmonella
24	185	8.5	286	2 Q7DHD3	Q7DHD3 serratia ma
25	185	8.5	286	2 Q799Y1	Q799Y1 plasmid ppv
26	185	8.5	286	2 Q7BVP8	Q7BVP8 plasmid pet
27	183	8.4	360	2 Q7AH57	Q7AH57 escherichia
28	183	8.4	360	2 Q8X684	Q8X684 escherichia
29	181	8.3	225	2 Q38212	Q38212 bacterioph
30	181	8.3	281	2 Q6QIV0	Q6QIV0 serratia ma
31	181	8.3	282	2 Q6QIV1	Q6QIV1 serratia ma
32	181	8.3	286	2 Q8KSD3	Q8KSD3 klebsiella
33	181	8.3	286	2 Q93A77	Q93A77 escherichia
34	181	8.3	286	2 Q93G13	Q93G13 klebsiella
35	181	8.3	286	2 Q6SQJ9	Q6SQJ9 streptococc
36	181	8.3	286	2 Q6UVM7	Q6UVM7 acinetobact
37	181	8.3	286	2 Q6W7J4	Q6W7J4 escherichia
38	181	8.3	286	2 Q8GA85	Q8GA85 escherichia
39	181	8.3	286	2 Q8EYX1	Q8EYX1 escherichia
40	181	8.3	286	2 Q9K582	Q9K582 klebsiella
41	181	8.3	286	2 Q9R771	Q9R771 klebsiella
42	181	8.3	286	2 Q9RMS4	Q9RMS4 escherichia
43	181	8.3	286	2 Q9RN48	Q9RN48 escherichia
44	181	8.3	286	2 Q9S679	Q9S679 escherichia
45	180	8.3	286	2 Q933Z8	Q933Z8 escherichia

ALIGNMENTS

RESULT 1
DGAL_ECOLI
ID DGAL_ECOLI STANDARD; PRT; 332 AA.
AC P02927; P17775;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE D-galactose-binding periplasmic protein precursor (GBP) (D-galactose/
D-glucose binding protein) (GGBP).
GN Name=mgIB; OrderedLocustNames=b2150, c2684, SF2235, S2364;
OS Escherichia coli,
OS Escherichia coli O6, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 623;
RN [1]_TaxID=562, 217992, 623;
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=92049246; PubMed=1719366;
RA Hogg R.W., Voelker C., von Carlowitz I.;
RT "Nucleotide sequence and analysis of the mgI operon of Escherichia
coli K12".
RL Mol. Gen. Genet. 229:453-459(1991).
RN [2]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF GLY-74.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=87286407; PubMed=3302609;
RA Scholle A., Vreemann J., Blank V., Nold A., Boos W., Manson M.D.;
RT "Sequence of the mgIB gene from Escherichia coli K12: comparison of
wild-type and mutant galactose chemoreceptors".
RL Mol. Gen. Genet. 208:247-253(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / BHB2600;
RA Richterich P., Leakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474 (1997).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic *Escherichia coli*.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
RN [6]
RP SEQUENCE OF 24-332.
RC SPECIES=E.coli;
RX MEDLINE=81168234; PubMed=7012152;
RA Mahoney W.C., Hogg R.W., Hermodson M.A.;
RT "The amino acid sequence of the D-galactose-binding protein from
Escherichia coli B/r.";
RL J. Biol. Chem. 256:4350-4356 (1981).
RN [7]
RP REVISIONS TO 179-180.
RC SPECIES=E.coli;
RA Mahoney W.C., Hogg R.W., Hermodson M.A.;
RL Submitted (NOV-1982) to the PIR data bank.
RN [8]
RP SEQUENCE OF 1-58 FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=83291030; PubMed=6885805;
RA Scripture J.B., Hogg R.W.;
RT "The nucleotide sequences defining the signal peptides of the
galactose-binding protein and the arabinose-binding protein.";
RL J. Biol. Chem. 258:10853-10855 (1983).
RN [9]
RP SEQUENCE OF 24-47.
RC SPECIES=E.coli; STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:1259-1313 (1997).
RN [10]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
through comparison with genomes of *Escherichia coli* K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [11]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA DOI=10.1128/IAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of *Shigella*
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
RN [12]

RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RC SPECIES=E.coli;
RX MEDLINE=83169767; PubMed=6340108;
RA Vyas N.K., Vyas M.N., Quicho F.A.;
RT "The 3-A resolution structure of a D-galactose-binding protein for
transport and chemotaxis in *Escherichia coli*.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:1792-1796 (1983).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RC SPECIES=E.coli;
RX MEDLINE=87258173; PubMed=3600760; DOI=10.1038/327635a0;
RA Vyas N.K., Vyas M.N., Quicho F.A.;
RT "A novel calcium binding site in the galactose-binding protein of
bacterial transport and chemotaxis.";
RL Nature 327:635-638 (1987).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RC SPECIES=E.coli;
RX MEDLINE=89058638; PubMed=3057628;
RA Vyas N.K., Vyas M.N., Quicho F.A.;
RT "Sugar and signal-transducer binding sites of the *Escherichia coli*
galactose chemoreceptor protein.";
RL Science 242:1290-1295 (1988).
CC -!- FUNCTION: This protein is involved in the active transport of
the two sugars by interacting with the trg chemoreceptor.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- DOMAIN: The calcium-binding site is structurally similar to that
of EF-hand proteins, but is in two parts, with the last calcium
ligand provided by Glu-228.
CC -!- SIMILARITY: Belongs to the bacterial extracellular solute-binding
protein family 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch.
CC -----
DR EMBL; M59444; AAA241169.1; -;
DR EMBL; U00007; AAA60523.1; -;
DR EMBL; X05646; CAA29132.1; -;
DR EMBL; U00096; AAC75211.1; -;
DR EMBL; AE016763; AAN81140.1; -;
DR EMBL; K00419; AAA24172.1; -;
DR EMBL; AE015239; AAN43756.1; -;
DR EMBL; AE016985; AAP17573.1; -;
DR PIR; A37277; JGECG.
DR PDB; 1GLG; X-ray; @=24-332.
DR PDB; 2GBP; X-ray; @=24-332.
DR SWISS-2DPAGE; P02927; COLI.
DR EC02DRASE; C029.8; 6TH EDITION.
DR EchoBASE; EB0588; -;
DR Ecogene; Egi0593; mgIB.
DR InterPro; IPR001761; PeriPlaBP/LacI.
DR Pfam; PF00532; PeriPla_BP_1; 1.
KW 3D-structure; Calcium-binding; Chemotaxis; Complete proteome;
KW Direct protein sequencing; Periplasmic; Signal; Sugar transport;
KW Transport.
FT SIGNAL 1 23
FT CHAIN 24 332
FT CA_BIND 157 165
FT CA_BIND 227 228
FT SITE 97 97
FT MUTAGEN 97 97
FT STRAND 26 32
FT TURN 35 36
FT HELIX 38 51
FT TURN 52 53
FT TURN 55 56
FT
FT D-galactose-binding periplasmic protein.
FT First part of site.
FT Second part of site.
FT Interacts with membrane-bound trg signal
FT transducer.
FT G->D: Improductive interaction with trg.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:10:32 ; Search time 802 Seconds
(without alignments)
6490.051 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 11920

Sequence: 1 gtttgacagcttatcatcg.....agttagcggaattgatctg 6729

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/can2_1/USPTO spool_p/US10649433/runat_02082005_164944_1513/app_query.fasta_1.6919
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CGN 1 1 977 @runat_02082005_164944_1513 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8639.5	72.5	1967	AB882793	Amino aci
2	4004	33.6	799	ABr43385	FLiPglu f
3	3273	27.1	2606	AD552217	pG1EN-EH3
4	3273	27.1	2606	AD552214	pG1EN-EH3
5	3086.5	25.9	2761	ADH11249	Vertebrat
6	3083.5	25.6	3614	ADH11258	Vertebrat
7	3062.5	25.7	766	ADH11236	Vertebrat
8	3058	25.7	1221	AAR52699	Sequence
9	3024	25.1	2732	ADH11251	Vertebrat
10	3004	24.9	1310	ADC22809	Human G p

C	11	3004	24.9	1310	7	ADH14282	Adh14282 Vector pc
	12	2946.5	24.7	864	7	ADH45227	Modified
	13	2902	24.3	1293	7	ADC22811	Human G p
	14	2902	24.3	1293	7	ADH14284	Vector pc
C	15	2901.5	24.1	2458	2	AAR07640	Deduced p
C	16	2901.5	24.1	2458	2	AAR04031	Full leng
C	17	2895.5	24.3	1277	2	AAR52701	Plasmid p
C	18	2890.5	24.0	2363	7	ADG15745	Novel vac
	19	2871	24.1	1277	2	AAR52700	Plasmid p
C	20	2842	23.6	1273	7	ADC22808	Human G p
C	21	2842	23.6	1273	7	ADH14281	Vector pc
	22	2824	23.7	1279	7	ADC22812	Human G p
	23	2824	23.7	1279	7	ADH14285	Vector pc
C	24	2819	23.4	1286	7	ADC22810	Human G p
C	25	2819	23.4	1286	7	ADH14283	Vector pc
C	26	2802	23.2	2050	2	AAR07641	Deduced s
	27	2773.5	23.3	1284	7	ADC22813	Human G p
	28	2773.5	23.3	1284	7	ADH14286	Vector pc
	29	2709	22.7	1088	2	AAR88636	Plasmid p
C	30	2707.5	22.5	2037	2	AAR04032	Full leng
	31	2675.5	22.4	1088	2	AAR88635	Plasmid p
	32	2626	22.0	2234	2	ADH11230	Vertebrat
	33	2618.5	22.0	1088	2	AAR88634	Plasmid p
	34	2617	22.0	763	6	ABR43387	FLiPbs f
	35	2592	21.7	2319	2	ADH11228	Vertebrat
	36	2586	21.7	1699	2	ADH11378	Vertebrat
	37	2486.5	20.9	740	4	AAB99360	Human H-R
	38	2486.5	20.9	740	4	AAB99357	Human H-R
	39	2486.5	20.9	740	5	ABBO6725	Human H-R
	40	2486.5	20.9	740	5	ABBO6728	Human H-R
	41	2486.5	20.9	740	5	ABBO6738	Human H-R
	42	2486.5	20.9	740	5	ABBO6735	Human H-R
	43	2486.5	20.9	740	5	ABBO6730	Human H-R
	44	2486.5	20.9	764	5	ABBO6740	Human H-R
	45	2486.5	20.9	764	5	ABBO6740	Human H-R

ALIGNMENTS

RESULT 1

AB882793

ID AB882793 standard; protein; 1967 AA.

XX AC AB882793;

XX AC

DT 18-MAR-2003 (first entry)

XX DE Amino acid sequence of plasmid pTrcCFRET3.

XX DE Fluorescent; fluorescence resonance energy transfer; FRET; tumour;

KW estrogen; tamoxifen; therapy; pTrcCFRET3.

XX OS Synthetic.

XX FH Key

FT Misc-difference 1. 1967

XX PN WO200290987-A2.

XX PD 14-NOV-2002.

XX PF 10-MAY-2002; 2002WO-GB002183.

XX PR 10-MAY-2001; 2001GB-00011459.

XX PA (ISIS-) ISIS INNOVATION LTD.

PI Fricker MD, Vaux DJT;

XX DR WPI; 2003-129191/12.

XX DR N-PSDB; AB223939.

XX

PT Probe, useful e.g. for medical diagnosis, and detection of pollutants in
 PT water systems and contaminants in foodstuffs, has target binding site
 PT moiety and mimic moiety attached to two fluorescent polypeptides, and
 PT linker.

XX Disclosure; Fig 5A-H; 50pp; English.

CC The invention relates to a probe (I) comprising a target binding site
 CC moiety attached to a first fluorescent polypeptide, a mimic moiety
 CC capable of binding to the target binding site moiety and attached to a
 CC second fluorescent polypeptide, and a linker connecting the two
 CC fluorescent polypeptides and which allows the distance between the
 CC fluorescent polypeptides to vary. (I), or a cell harbouring (I), or a
 CC polynucleotide which encodes (I), or a sensor (V) comprising (I), a light
 CC source which is capable of exciting the probe, and a detecting which is
 CC capable of measuring the amount of fluorescence resonance energy transfer
 CC (FRET) from the probe are useful for detecting the presence or absence of
 CC a target substance in a test sample, and for identifying an inhibitor of
 CC binding between two substances, where the two substances would bind to
 CC each other in the absence of an inhibitor (see AB223939 for a description
 CC of the varied uses of the probe of the invention). The present sequence
 CC represents the amino acid sequence of plasmid pTrcCFRET3

XX Sequence 1967 AA;

SQ Alignment Scores:

Pred. No.: 0 Length: 1967
 Score: 8639.50 Matches: 1701
 Percent Similarity: 81.30% Conservative: 21
 Best Local Similarity: 80.31% Mismatches: 76
 Query Match: 72.48% Indels: 320
 DB: 6 Gaps: 8

US-10-649-433C-1 (1-6729) x ABB82793 (1-1967)

QY 413 ATGTGAGCAAGGCGAGAGCTGTTACCGGGGTGGTCCCATCTCGTGGAGTGGAC 472
 DB MetValSerLysGlyGluLeuLeuPheThrGlyValValProIleLeuValGluLeuAsp 23
 QY 473 GCGCAGCTAAACGGCCACAAGTTACGCTGTCGCGCGAGGGCGAGGGGATGCCACCTAC 532
 DB 24 GlyAspValAsnGlyHisLysPheSerValSerGlyGlyGlyGluGlyAlaThrTyr 43
 QY 533 GCGAAGCTGACCTCAAGTTGATCTGACACCGGCAAGCTGCCGTGCTGGCCACC 592
 DB 44 GlyLysLeuThrLeuLysPheIleCysThrThrGlyLysLeuProValProTrpProThr 63
 QY 593 CTCGTGACCACCTTCGGCTACGGCTGACGTGCTTCCGCGCTACCCCGACACATGAAG 652
 DB 64 LeuValThrLeuThrTrpGlyValGlnCysPheSerArgTyrProAspHisMetLys 83
 QY 653 CAGCAGCACTTCTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC 712
 DB 84 GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 103
 QY 713 TTCAGGAGCAGCGCACTACAGACCGCGCGAGGTGAAGTTCGAGGGCGACCCCTG 772
 DB 104 PheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 123
 QY 773 GTGAACCGCATCGAGTGAAGGGCATCGACTTCAAGGAGGAGCGCAACATCTCGGGCAC 832
 DB 124 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 143
 QY 833 AAGCTGGAGTACAATACACAGCCACCAACGTCTATATCATGGCCGACAGCAGAGAAC 892
 DB 144 LysLeuGluTyrAsnTyrIleSerHisAsnValTyrIleThrAlaAspLysGlnLysAsn 163
 QY 893 GGCATCAGGTGAATTCAGATCCGCCACACATCGAGGAGCGGAGGTGCGAGTCCGCC 952
 DB 164 GlyIleLysAlaAsnPheLysIleArgHisAsnIleGluAspGlySerValGlnLeuAla 183
 QY 953 GACCACTACGACAGAACACCCCATCGCGACCGCGCTGCTGCTGCCCGCAACACAC 1012

DB 184 AspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsnHis 203
 QY 1013 TACCTGAGTACCACTCGCCCTGAGCAAAACCCCAACGAGACGCGATCACATGTC 1072
 DB TyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMetVal 223
 QY 1073 CTGCTGGAGTTCGTGACCGCGCC-----GGGATCACTCTC 1108
 DB 224 LeuLeuGluPheValThrAlaAlaLeuGlnSerSerGlyGlyGlyGlyGlyMet 243
 QY 1109 GGCATGGACGAG-----CTGTACAAGACTAGTGTCTGATCACTCGCATT 1150
 DB 244 GlyGlySerHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 263
 QY 1151 GGTGTAACTTATAAGTACGACGATAACTTTATGCTGTAGTGCAGGCTATTGAG 1210
 DB 264 GlyArgAspLeuTyrAspAspAsp----- 272
 QY 1211 CAAGATGGAAGCGCGCGCCAGATGTTTCAGTGTGATGAATGATTCAGATGACAC 1270
 DB 272 ----- 272
 QY 1271 TCCAAGCAGAACGATCAGATCGACTATTGCTGGCCAGGGGTCAAGGCTCGCCATC 1330
 DB 272 ----- 272
 QY 1331 AACCTGTTGACCGCGCAGCTGCGGTACGCTGATGTTGAGAAAGCGCGTGGGCAAAACGTG 1390
 DB 272 ----- 272
 QY 1391 CCGTGGTTCCTTCAACAAGAACCGCTCTCGTAAGGCGCTGGATAGCTACGACAAAGCC 1450
 DB 272 ----- 272
 QY 1451 TACTACGTTGGCACTGACTCAAAAGAGTCGGCATTTATTCAAGGCGATTTGATGCTATA 1510
 DB 273 -----Lys 273
 QY 1511 CACTGGGCGGCAATCAGGGTTGGATCTGAACAAAGCGTCAAGTTCAGTTCGTCTG 1570
 DB 274 His-----ArgTrpIleArg 278
 QY 1571 CTGAAGGTGAACCGGGCCATCCGGATCGAGACGACGACCTACCTACGTGATTAAGAA 1630
 DB 279 ProArgGlySerSerGly----- 284
 QY 1631 TTGAACGATAAAGGCATCAAAACTGAACAGTTACAGTTAGATACCGCAATGTGGACACC 1690
 DB 284 ----- 284
 QY 1691 GCTCAGGCGAAGATAAGATGAGCGCTGGCTGTCTGGCCCGAAGCGCAACAAATCGAA 1750
 DB 284 ----- 284
 QY 1751 GTGGTTATGCCAACACGATCGCATGGCGGATGGCGGTTGAAGCGCTGAAAGCACAC 1810
 DB 284 ----- 284
 QY 1811 AACAACTCCAGCATTCCTGGTGTGGCGCTCGATGCTGCCAGAACGCGCTGCGCGTGTG 1870
 DB 284 ----- 284
 QY 1871 AAATCCGCTGACCTGGCGGCGACCGTACTGAACGATGCTAACACCGGCGGAAGCGACC 1930
 DB 285 -----GlyGlyGlySerGly----- 289
 QY 1931 TTTGATCTGCGGAAAAACCTCGCGCATGTAAGTGGCTGATGCGACCACTGGA 1990
 DB 290 -----GlyGlyGlySerGlyGly----- 296
 QY 1991 ATCGACAACAAAGGTGTCGCGTACCTTATGTTGGCGTAGATAAAGACAACCTGCTGAA 2050
 DB 296 ----- 296

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: August 2, 2005, 18:20:38 ; Search time 160 Seconds
(without alignments)
6278.919 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 11920
Sequence: 1 gtttagacgattcatcg.....agttagcggaattgatctg 6729

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/USPTO spool_p/US10649433/runat_02082005_164946_1547/app_query.fasta_1.6919
-DB=issued_Patents_AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CGN 1 1 178 @runat_02082005_164946_1547 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	3004	24.9	1310	4	US-09-170-496D-290
C 2	3004	24.9	1310	4	US-09-364-425B-55
C 3	2902	24.3	1293	4	US-09-170-496D-292
C 4	2902	24.3	1293	4	US-09-364-425B-57
C 5	2842	23.6	1273	4	US-09-170-496D-289
C 6	2842	23.6	1273	4	US-09-364-425B-54
C 7	2824	23.7	1279	4	US-09-170-496D-293
C 8	2824	23.7	1279	4	US-09-364-425B-58
C 9	2819	23.4	1286	4	US-09-170-496D-291
C 10	2819	23.4	1286	4	US-09-364-425B-56
C 11	2773.5	23.3	1284	4	US-09-170-496D-294
C 12	2773.5	23.3	1284	4	US-09-364-425B-59

13	2449	20.5	1056	4	US-09-513-783A-32	Sequence 32, Appl
14	2449	20.5	1056	4	US-09-430-656-32	Sequence 32, Appl
15	2384.5	20.0	541	4	US-09-513-783A-34	Sequence 34, Appl
16	2384.5	20.0	541	4	US-09-430-656-34	Sequence 34, Appl
17	2360.5	19.8	642	2	US-08-818-253-6	Sequence 6, Appl
18	2360.5	19.8	642	3	US-08-818-253-6	Sequence 6, Appl
19	2360.5	19.8	652	2	US-08-818-253-4	Sequence 4, Appl
20	2360.5	19.8	652	3	US-08-818-253-4	Sequence 4, Appl
21	2359.5	19.8	642	2	US-08-818-253-2	Sequence 2, Appl
22	2359.5	19.8	642	3	US-08-818-253-2	Sequence 2, Appl
23	2337.5	19.6	656	2	US-08-818-253-8	Sequence 8, Appl
24	2337.5	19.6	656	3	US-08-818-253-8	Sequence 8, Appl
25	2216.5	18.6	1610	4	US-09-513-783A-22	Sequence 22, Appl
26	2216.5	18.6	1610	4	US-09-430-656-22	Sequence 22, Appl
27	1774	14.9	360	5	PCT-US93-01676A-2	Sequence 4, Appl
28	1771	14.9	370	5	PCT-US93-01676A-4	Sequence 4, Appl
29	1768	14.8	360	3	US-09-025-769B-293	Sequence 293, App
30	1768	14.8	360	4	US-09-490-070A-293	Sequence 293, App
31	1768	14.8	360	4	US-09-490-153-293	Sequence 293, App
32	1768	14.8	360	4	US-09-490-324-293	Sequence 293, App
33	1765	14.8	1049	3	US-09-522-666-6	Sequence 6, Appl
34	1759	14.8	441	3	US-09-522-666-4	Sequence 4, Appl
35	1759	14.8	1024	3	US-09-522-666-2	Sequence 2, Appl
36	1754	14.7	360	1	US-08-456-923-21	Sequence 21, Appl
37	1569	13.2	319	3	US-09-044-796A-13	Sequence 13, Appl
38	1569	13.2	319	4	US-09-725-460A-13	Sequence 13, Appl
39	1510	12.7	353	4	US-09-489-039A-8090	Sequence 8090, Ap
40	1503	12.6	299	3	US-09-025-769B-285	Sequence 285, App
41	1503	12.6	299	3	US-09-025-769B-298	Sequence 298, App
42	1503	12.6	299	3	US-09-025-769B-300	Sequence 300, App
43	1503	12.6	299	4	US-09-490-070A-285	Sequence 285, App
44	1503	12.6	299	4	US-09-490-070A-298	Sequence 298, App
45	1503	12.6	299	4	US-09-490-070A-300	Sequence 300, App

ALIGNMENTS

RESULT 1
US-09-170-496D-290
; Sequence 290, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 290
; LENGTH: 1310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-290

Alignment Scores:
Pred. No.: 2.15e-238 Length: 1310
Score: 3004.00 Matches: 694
Percent Similarity: 65.51% Conservative: 45
Best Local Similarity: 61.52% Mismatches: 165
Query Match: 24.91% Indels: 224
DB: 4 Gaps: 42

US-10-649-433C-1 (1-6729) x US-09-170-496D-290 (1-1310)

QY 6352 TCTGATCTTGGCAACGACATCGCGGAGGACGATCCCTCATTGACATTCGATGG 6293
:::|||||:::|||||
Db 84 ThrAspGluIrrpGlu-----GlnTrp-----trp 91
QY 6292 -----TTTGTGTAACCGGACATCGCGCTTCGCTCGCTATCGGCT 6239

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: August 2, 2005, 18:23:19 ; Search time 801 Seconds
(without alignments)
6545.998 Million cell updates/sec
Title: US-10-649-433C-1
Perfect score: 11920
Sequence: 1 gtttagcagcttatcatcga.....agtttagcgcgaattgatctg 6729
Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 1745140 seqs, 389608008 residues
Total number of hits satisfying chosen parameters: 3490280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10649433/runat_02082005_164946_1571/app_query.fasta_1.6919
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -LOCAL=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649433 @CGN 1 1 964 runat 02082005 164946_1571
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DBLEXT=7

Database :
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pcp.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pcp.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8639.5	72.5	1967	16	US-10-477-044-1 Sequence 1, Appli
c 2	3004	24.9	1310	14	US-10-251-385-290 Sequence 290, App
c 3	3004	24.9	1310	18	US-10-668-035-55 Sequence 55, Appl
4	2902	24.3	1293	14	US-10-251-385-292 Sequence 292, App
5	2902	24.3	1293	18	US-10-668-035-57 Sequence 57, Appl
c 6	2842	23.6	1273	14	US-10-251-385-289 Sequence 289, App
c 7	2842	23.6	1273	18	US-10-668-035-54 Sequence 54, Appl
8	2824	23.7	1279	14	US-10-251-385-293 Sequence 293, App
9	2824	23.7	1279	18	US-10-668-035-58 Sequence 58, Appl
c 10	2819	23.4	1286	14	US-10-251-385-291 Sequence 291, App
c 11	2819	23.4	1286	18	US-10-668-035-56 Sequence 56, Appl
12	2773.5	23.3	1284	14	US-10-251-385-284 Sequence 284, App
13	2773.5	23.3	1284	18	US-10-668-035-59 Sequence 59, Appl
14	2486.5	20.9	740	15	US-10-344-404-12 Sequence 12, Appl
15	2486.5	20.9	740	15	US-10-344-404-27 Sequence 27, Appl
16	2486.5	20.9	764	15	US-10-344-404-34 Sequence 34, Appl
17	2463.5	20.7	740	15	US-10-344-404-23 Sequence 23, Appl
18	2449	20.5	1056	14	US-10-100-957A-32 Sequence 32, Appl
19	2416.5	20.3	734	15	US-10-344-404-31 Sequence 31, Appl
20	2416.5	20.3	758	15	US-10-344-404-36 Sequence 36, Appl
21	2413	20.2	745	15	US-10-344-404-19 Sequence 19, Appl
22	2395	20.1	823	15	US-10-344-404-42 Sequence 42, Appl
23	2387.5	20.0	762	15	US-10-344-404-40 Sequence 40, Appl
24	2386.5	20.0	763	15	US-10-344-404-38 Sequence 38, Appl
25	2384.5	20.0	541	14	US-10-100-957A-34 Sequence 34, Appl
26	2371	19.9	723	17	US-10-732-923-14279 Sequence 14279, A
27	2360.5	19.8	642	9	US-09-554-000-6 Sequence 6, Appli
28	2360.5	19.8	652	9	US-09-554-000-4 Sequence 4, Appli
29	2359.5	19.8	642	9	US-09-554-000-2 Sequence 2, Appli
30	2359	19.8	803	16	US-10-634-740-7 Sequence 7, Appli
31	2350	19.7	635	16	US-10-634-740-5 Sequence 5, Appli
32	2337.5	19.6	656	9	US-09-554-000-8 Sequence 8, Appli
33	2216.5	18.6	1610	14	US-10-100-957A-22 Sequence 22, Appl
34	2201.5	18.5	607	10	US-09-293-670-55 Sequence 55, Appl
35	2192.5	18.4	477	10	US-09-293-670-54 Sequence 54, Appl
36	1788	15.0	363	10	US-09-987-763-29 Sequence 29, Appl
37	1771	14.9	372	16	US-10-469-881-3 Sequence 3, Appli
38	1765	14.8	360	13	US-10-197-053-1 Sequence 1, Appli
39	1758	14.7	360	13	US-10-197-053-2 Sequence 2, Appli
40	1579	13.2	332	13	US-10-197-053-6 Sequence 6, Appli
41	1579	13.2	332	17	US-10-893-671-82 Sequence 82, Appl
42	1578	13.2	309	16	US-10-840-796-1 Sequence 1, Appli
43	1569	13.2	319	9	US-09-725-460A-13 Sequence 13, Appl
44	1569	13.2	319	20	US-11-004-853-15 Sequence 15, Appl
45	1486	12.5	1213	15	US-10-381-882-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-10-477-044-1
; Sequence 1, Application US/10477044
; Publication No. US20040265902A1
; GENERAL INFORMATION:
; APPLICANT: Fricker, Mark
; APPLICANT: Vaux, David
; TITLE OF INVENTION: Universal Fluorescent Sensors
; FILE REFERENCE: HO-P02985US0
; CURRENT APPLICATION NUMBER: US/10/477,044
; CURRENT FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: PCT/GB02/02183
; PRIOR FILING DATE: 2002-02-10
; PRIOR APPLICATION NUMBER: GB 0111459.4
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1967
; TYPE: PRT

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 19:37:23 ; Search time 27846 Seconds
(without alignments)
11709.227 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gttgacagcttcatcg.....agttagcgcgaattgatctg 6729

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_scs.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4344.4	64.6	5271	12	CVU89936
2	3855	57.3	4806	12	AF050464
3	3854.2	57.3	4205	6	AX576406
4	3853.4	57.3	5013	6	AX264513
5	3853.2	57.3	4203	6	AX576404
6	3853.2	57.3	4204	6	AX576405
7	3853.2	57.3	4217	6	AX576408
8	3853.2	57.3	4218	6	AX576409
9	3852.8	57.3	4216	6	AX576407
10	3852.2	57.2	4176	6	AI3038
11	3852.2	57.2	4176	12	SYNTRC99A
12	3852.2	57.2	4239	6	A29289
13	3851.6	57.2	4476	6	ARI02989
14	3851.6	57.2	4476	6	ARI172783
15	3851.4	57.2	4214	6	BD273391
16	3851.4	57.2	4214	6	AX027488
17	3850.6	57.2	4176	12	XXU13872
18	3824.8	56.8	4232	12	SYNLACIQ
19	3824.8	56.8	4735	12	SYNBETALAC

c	20	3391.8	50.4	6862	6	AI0361	AI0361 Artificial
c	21	3382.4	50.3	4733	12	XXU19585	UI9585 Cloning vec
c	22	3375.2	50.2	4361	12	AY327137	AY327137 Expressio
c	23	3375.2	50.2	5189	12	AY327139	AY327139 Expressio
c	24	3344.8	49.7	5024	6	CQ846977	CQ846977 Sequence
c	25	3343.2	49.7	4803	12	AY243506	AY243506 Cloning v
c	26	3341.2	49.7	6464	6	BD263349	BD263349 Compositi
c	27	3304.4	49.1	4770	12	AF333022	AF333022 Expressio
c	28	3297.4	49.0	5885	1	STAFRA	L42764 Staphylococ
c	29	3230.2	48.0	5398	6	AX824326	AX824326 Sequence
c	30	2889.8	42.9	8811	12	AY569779	AY569779 Cloning v
c	31	2889.8	42.9	12404	12	AY569780	AY569780 Cloning v
c	32	2864.8	42.6	4109	12	PSPORT1	UI2390 Cloning vec
c	33	2864.8	42.6	5957	6	BD263353	BD263353 Compositi
c	34	2864.8	42.6	5957	6	BD263354	BD263354 Compositi
c	35	2863.2	42.6	5451	6	AX771237	AX771237 Sequence
c	36	2859.6	42.5	4932	12	SYNECOPGEX	M21676 Plasmid pGE
c	37	2859.6	42.5	4968	12	SYNGLUTRAN	M97937 Plasmid pGE
c	38	2859.6	42.5	4984	12	CVU78874	U78874 pGEX-6P-3 c
c	39	2859.6	42.5	4984	12	CVU78872	U78872 Cloning vec
c	40	2859.6	42.5	4985	12	CVU78873	U78873 pGEX-6P-2 c
c	41	2859.6	42.5	5487	12	FGH223813	A0223813 Synthetic
c	42	2858	42.5	4506	12	AB116367	AB116367 Expressio
c	43	2858	42.5	4557	6	EI3740	EI3740 DNA Sequence
c	44	2858	42.5	4557	6	AR373251	AR373251 Sequence
c	45	2858	42.5	4947	12	XXU13849	UI13849 pGEX-1 Lamb

ALIGNMENTS

RESULT 1
CVU89936
LOCUS
DEFINITION Cloning vector p6xHis-GFP, complete sequence.
ACCESSION U89936
VERSION U89936.1 GI:2190721
KEYWORDS
SOURCE unidentified cloning vector
ORGANISM unidentified cloning vector
REFERENCE 1 (bases 1 to 5271)
AUTHORS Kitts, P.
TITLE p6xHis-GFP, complete sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5271)
AUTHORS Holtz, A. and Lou, Y.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-1997) CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230, USA
COMMENT This vector can be obtained from CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230, USA. To place an order, call (415) 424-8222 or (800) 662-2566, extension 1. International customers, please contact your local distributor or subsidiary. For technical information, call (415) 424-8222 or (800) 662-2566, extension 3.
This sequence has been compiled from information in the sequence databases, published literature and other sources, together with partial sequences obtained by CLONTECH. If you suspect there is an error in this sequence, please contact CLONTECH's Technical Support Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.

FEATURES
source Location/Qualifiers
1..5271
/organism="unidentified cloning vector"
/mol_type="genomic DNA"
/db_xref="taxon:45196"

ORIGIN
Query Match 64.6%; Score 4344.4; DB 12; Length 5271;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 4582; Conservative 0; Mismatches 81; Indels 159; Gaps 3;

QY 2067 CAGTAAAGGAGAACTTTTCACTGGAGTGTGCCAATCTCTGTTGAAATTAGATGGT 2126
DB |||||
450 CTAGCAAGGAGAGAACTTTTCACTGGAGTGTGCCAATCTCTGTTGAAATTAGATGGT 509
QY 2127 ATGTTAATGGGCACAAATTTTCTGTCACTGGAGAGGGTGAAGGTGATGCAACATACGGAA 2186
DB |||||
510 ATGTTAATGGGCACAAATTTTCTGTCACTGGAGAGGGTGAAGGTGATGCAACATACGGAA 569
QY 2187 AACTTACCTTAAATTTATTTGCACTACTGGAATACTACTGTTTCCATGGCCAACTTTG 2246
DB |||||
570 AACTTACCTTAAATTTATTTGCACTACTGGAATACTACTGTTTCCATGGCCAACTTTG 629
QY 2247 TCACCTACTCTTCTTTATGGTGTCAATGCTTTTCCGTTATCCGATCATATGAAACGGC 2306
DB |||||
630 TCACCTACTCTTCTTTATGGTGTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGC 689
QY 2307 ATGACTTTTCAAGAGTGCATGCCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCA 2366
DB |||||
690 ATGACTTTTCAAGAGTGCATGCCCCGAAGGTTATGTACAGGAAGAATACTATATTTTCA 749
QY 2367 AAGATGACGGGAACCTACAGAGCGGTGCTGAAGTCAAGTTTGAAGGTGATACCCCTGTTA 2426
DB |||||
750 AAGATGACGGGAACCTACAGAGCACGTGCTGAAGTCAAGTTTGAAGGTGATACCCCTGTTA 809
QY 2427 ATCGTATCGAGTTAAAGGTATGATTTTAAAGAGATGGAACATTTCTCGACACAAAC 2486
DB |||||
810 ATAGAATCGAGTTAAAGGTATGATTTTAAAGAGATGGAACATTTCTTGACACAAAT 869
QY 2487 TCGAGTCAAACTATATACTCACACAATGTATACATCACGGCAGACAAACAAAGAAATGAA 2546
DB |||||
870 TGGATAACAATACTAATCTCACACATGTATACATCGCAGACAAACAAAGAAATGAA 929
QY 2547 TCAAAGCTAACTTCAAATTCGCCACAACTGAAGATGGATTCGGTCAACTAGCAGACC 2606
DB |||||
930 TCAAAGTTAACTTCAAATTTAGACACAACTGAAGATGGAAGCTTCAACTAGCAGACC 989
QY 2607 ATTATCAACAAATCTCCAAATGGCGATGGCCGTCTCTTTACGACACCAATACC 2666
DB |||||
990 ATTATCAACAAATCTCCAAATGGCGATGGCCGTCTCTTTACGACACCAATACC 1049
QY 2667 TGTGACACAACTCTGCCCTTTGAAAGATCCCAAGAAAGCGTGACACACATGGTCTTTC 2726
DB |||||
1050 TGTCCACACAACTCTGCCCTTTGAAAGATCCCAAGAAAGAGACACACATGGTCTTTC 1109
QY 2727 TTGAGTTGTAACTGCTGCGGATTAACATGGCATGGATGAGTCTTACAAATAAAG- 2785
DB |||||
1110 TTGAGTTGTAACTGCTGCGGATTAACATGGCATGGATGAGTCTTACAAATAAATGT 1169
QY 2786 -----CTTACGTAGAACAAACTCA----- 2806
DB |||||
1170 CCAGACTTCCAAATTGACACTAAGTGTCCGAACAAATTACTAAATCTCAGGGTTCCTGGT 1229
QY 2807 ----- 2806
DB |||||
1230 TAAATTCAGGCTGAGATATTTATATATTTATAGATTCATTAATACTGATGATAAT 1289
QY 2807 -----TCTCAGAAAGAGATCTGAATAGCGCCGTCG 2836
DB |||||
1290 TTAATTGATGTTATTGATAGAGTTATTTTCTTAATTAACACGGCTACTTTGGAGTGTATTCT 1349
QY 2837 ACCATCATCATCATCATGATGAGTTTAAACGGTCTC----- 2873
DB |||||
1350 TAAITCTATATTAATTAACATTTGATTTGCTCTAAAGCGGCGCGCGCGCGCG 1409
QY 2874 -----CAGCTTGGCTGTTTGGCGGATGAGAAAGATTTTTCAGCCTGATACAGATTAAA 2927
DB |||||
1410 AATTGGAAGCTTGGCTGTTTGGCGGATGAGAAAGATTTTCAGCCTGATACAGATTAAA 1469
QY 2928 TCAGAACGAGAGCGGTCTGATAAAACAGAAATTTGCTGGCGGAGTAGCGGTGTC 2987
DB |||||
1470 TCAGAACGAGAGCGGTCTGATAAAACAGAAATTTGCTGGCGGAGTAGCGGTGTC 1529
QY 2988 CCACCTGACCCCATGCCAACTCAGAAAGTGAAACCGCGTAGCGCGGATGTGTGGG 3047

DB 1530 CCACCTGACCCCATGCCGAATCAGAAAGTGAAACCGCGTAGCGCGATGGTAGTGTGGG 1589
QY 3048 TCTCCCATCGAGAGTAGGGAACCTGCAGGCATCAATAAAGCAAGGCTCAGTCGAA 3107
DB 1590 TCTCCCATCGAGAGTAGGGAACCTGCAGGCATCAATAAAGCAAGGCTCAGTCGAA 1649
QY 3108 AGACTGGGCTTTGCTGTTTATCTGTTTGTGCGGTGAACGCTCTCTCTGAGTAGGACAAA 3167
DB 1650 AGACTGGGCTTTGCTGTTTATCTGTTTGTGCGGTGAACGCTCTCTCTGAGTAGGACAAA 1709
QY 3168 TCCGCGGGAGCGGATTTGAACGTTTGGAAAGCAACGCGCGAGGTCGGCGGACGAGCG 3227
DB 1710 TCCGCGGGAGCGGATTTGAACGTTTGGAAAGCAACGCGCGAGGTCGGCGGACGAGCG 1769
QY 3228 CCGGCCATAAACTGCCAGGCATCAAAATTAAGCAGAGGCCATCTCTGACGAGTGCCTTTT 3287
DB 1770 CCGGCCATAAACTGCCAGGCATCAAAATTAAGCAGAGGCCATCTCTGACGAGTGCCTTTT 1829
QY 3288 TCGGTTTCTACAAACTCTTTTGTGTTTATTTTCTAAATACATTCAAATATGATATCCGCTC 3347
DB 1830 TCGGTTTCTACAAACTCTTTTGTGTTTATTTTCTAAATACATTCAAATATGATATCCGCTC 1889
QY 3348 ATGAGACAAATAACCTCTGATAAATCTTCAATAATATTGAAAGCAAGAGTATGAGTAT 3407
DB 1890 ATGAGACAAATAACCTCTGATAAATCTTCAATAATATTGAAAGCAAGAGTATGAGTAT 1949
QY 3408 CAACATTTCCGTGTCGCCCTTATTTCCCTTTTTCGGCATTTTCCCTTCTCTGTTTTCCT 3467
DB 1950 CAACATTTCCGTGTCGCCCTTATTTCCCTTTTTCGGCATTTTCCCTTCTCTGTTTTCCT 2009
QY 3468 CACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAAGTCAAGTGGGTGACGAGTGGGT 3527
DB 2010 CACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAAGTCAAGTGGGTGACGAGTGGGT 2069
QY 3528 TACATCGAACTGGATCTCAACAGCGGTAAAGATCTTGAAGTGGTTCGGCCGCAAGAAAGT 3587
DB 2070 TACATCGAACTGGATCTCAACAGCGGTAAAGATCTTGAAGTGGTTCGGCCGCAAGAAAGT 2129
QY 3588 TTTCCAAATGATGAGCACCTTTTAAAGTCTGTATGTGCGCGGTATTAATCCGCTGTTGAC 3647
DB 2130 TTTCCAAATGATGAGCACCTTTTAAAGTCTGTATGTGCGCGGTATTAATCCGCTGTTGAC 2189
QY 3648 GCCGGCAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAAATGACTTGGTTCAGTAC 3707
DB 2190 GCCGGCAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAAATGACTTGGTTCAGTAC 2249
QY 3708 TCACCACTCAGAAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATTAATGAGTGTCT 3767
DB 2250 TCACCACTCAGAAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATTAATGAGTGTCT 2309
QY 3768 GCCATAACCATGAGTATAACACTGCGGCCAACTTACTTCTTGACAAACGATCGGAGACCG 3827
DB 2310 GCCATAACCATGAGTATAACACTGCGGCCAACTTACTTCTTGACAAACGATCGGAGACCG 2369
QY 3828 AAGGAGCTTAAACGCTTTTTCGCAACAATGGGGGATCATGTAACTCGCCTTGTATGCTGG 3887
DB 2370 AAGGAGCTTAAACGCTTTTTCGCAACAATGGGGGATCATGTAACTCGCCTTGTATGCTGG 2429
QY 3888 GAAACGGAGCTGAATTAAGCCATACCAAAACGACGAGCGTGACACACCATGCTGTAGCA 3947
DB 2430 GAAACGGAGCTGAATTAAGCCATACCAAAACGACGAGCGTGACACACCATGCTGTAGCA 2489
QY 3948 ATGGCAACACGTTGCGCAAACTAATTAACCTGGGAACTACTTACTCTAGCTTCCCGGAA 4007
DB 2490 ATGGCAACACGTTGCGCAAACTAATTAACCTGGGAACTACTTACTCTAGCTTCCCGGAA 2549
QY 4008 CAATTAATAGACTCGAGCGGATTAAGTTCAGGACCACTTCTCGCTCGGCCCTT 4067
DB 2550 CAATTAATAGACTCGAGCGGATTAAGTTCAGGACCACTTCTCGCTCGGCCCTT 2609
QY 4068 CCGGCTGCTGTTTATTTGCTGATAAAATCTGGAGCGGTGAGCGGTGCTCGCGGTATC 4127

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 19:20:28 ; Search time 3192 Seconds
(without alignments)
12479.304 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttagacgttcatcatcg.....agtttagcggaattgatctg 6729

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4119.8	61.2	5903	8	ABZ23939 Nucleotid
2	3855	57.3	5027	2	AAX78865 Human tis
3	3855	57.3	5069	2	AAX78871 Human tis
4	3854.2	57.3	4205	6	ABQ74927 Avidity I
5	3853.4	57.3	4816	4	ABQ7799 pTRCHISB
6	3853.4	57.3	4816	12	ADf72546 Mouse Adi
7	3853.4	57.3	4816	12	ADG14708 Globular
8	3853.4	57.3	5013	9	ACA63355 E. coli D
9	3853.4	57.3	5013	10	AAD59423 E. coli pu
10	3853.4	57.3	5013	10	ADE86127 E. coli p
11	3853.4	57.3	5104	4	AAD07798 pTRCHISB
12	3853.4	57.3	5104	12	ADf72545
13	3853.2	57.3	4203	6	ABQ74925 Avidity I
14	3853.2	57.3	4204	6	ABQ74926 Avidity I
15	3853.2	57.3	4217	6	ABQ74929 Avidity I
16	3853.2	57.3	4218	6	ABQ74930 Avidity I
17	3852.8	57.3	4216	6	ABQ74928 Avidity I
18	3852.2	57.2	5926	2	AAX32978 Tn7 donor
19	3852.2	57.2	5926	2	AAX32977 Tn7 donor
20	3852.2	57.2	5926	6	AAD45060 pEM delta

C	21	3852.2	57.2	5926	6	AAD45059
C	22	3852.2	57.2	5926	12	ADG46817
	23	3851.8	57.2	5069	2	AAX78874
	24	3851.8	57.2	5099	2	AAX78896
	25	3851.8	57.2	5132	2	AAX78883
	26	3851.8	57.2	5324	2	AAX78890
	27	3851.8	57.2	5342	2	AAX78893
	28	3851.4	57.2	4214	3	AAX59260
	29	3851.4	57.2	4214	3	AAX59260 Nucleotid
	30	3850.6	57.2	4176	1	AAX62632 Vector pE
	31	3850.6	57.2	4176	1	AAN90709 Sequence
	32	3850.6	57.2	4176	12	AD136476 Plasmid C
C	33	3841.4	57.1	5103	12	ADG46818 pEM DNA (
	34	3727.8	55.4	4189	2	ADG14707 Mouse ORG
	35	3542	52.6	5161	10	AAQ05397 Secretion
	36	3344.8	49.7	5024	9	ADH45226 Modified
C	37	3344.8	49.7	5024	9	ACF06051 Plasmid p
	38	3341.2	49.7	6464	3	ADL18668 Plasmid p
	39	3341.2	49.7	6464	10	ABZ58765 Destinati
C	40	3241.8	48.2	7174	1	AAN91062 Sequence
	41	3230.2	48.0	5398	10	ADP50156 Plasmid G
C	42	2864.8	42.6	5957	3	AAC55467 Destinati
	43	2864.8	42.6	5957	3	AAC55464 Destinati
	44	2863.2	42.6	5451	9	ADA41729 Plasmid p
	45	2858	42.5	4557	2	AAT90491 Vector pl

ALIGNMENTS

RESULT 1

ABZ23939

ID ABZ23939 standard; DNA; 5903 BP.

AC ABZ23939;

DT 18-MAR-2003 (first entry)

DE Nucleotide sequence of plasmid pTRCpFRET3.

XX Fluorescent; fluorescence resonance energy transfer; FRET; tumour;

XX estrogen; tamoxifen; therapy; pTRCpFRET3; ds.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 2..5902

FT /*tag= a

FT /note= "contains internal stop codons"

XX WO200290987-A2.

XX 14-NOV-2002.

XX 10-MAY-2002; 2002WO-GB002183.

XX 10-MAY-2001; 2001GB-00011459.

XX (ISIS-) ISIS INNOVATION LTD.

XX Fricker MD, Vaux DJT;

XX WPI, 2003-129191/12.

XX P-PSDB; ABB82793.

XX Probe, useful e.g. for medical diagnosis, and detection of pollutants in

XX water systems and contaminants in foodstuffs, has target binding site

XX moiety and mimic moiety attached to two fluorescent polypeptides, and

XX linker.

XX Disclosure; Fig 5A-H; 50pp; English.

XX The invention relates to a probe (I) comprising a target binding site

CC

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 09:18:00 ; Search time 992 Seconds
(without alignments)
11099.300 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttagacgttatcatcga.....agttagcggaattgatctg 6729

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A-COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5B-COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A-COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B-COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/PCTUS-COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3853.4	57.3	5013	US-09-183-188B-5	Sequence 5, Appli
2	3852.2	57.2	5226	US-09-027-169-3	Sequence 3, Appli
3	3852.2	57.2	5226	US-09-027-169-4	Sequence 4, Appli
4	3851.6	57.2	4476	US-08-801-344-2	Sequence 2, Appli
5	3851.6	57.2	4476	US-09-498-599-2	Sequence 2, Appli
6	2858	42.5	4557	US-08-778-717-5	Sequence 5, Appli
7	2729	40.6	4357	US-09-795-872-10	Sequence 10, Appli
8	2598.8	38.6	4593	US-08-801-344-1	Sequence 1, Appli
9	2598.8	38.6	4593	US-09-498-599-1	Sequence 1, Appli
10	2596	38.6	6312	US-08-531-601-3	Sequence 3, Appli
11	2596	38.6	6312	US-08-859-032-3	Sequence 3, Appli
12	2591.8	38.5	5238	US-03-363-1	Patent No. 5453363
13	2591.8	38.5	5238	US-03-363-1	Patent No. 5453363
14	2483.6	36.9	4410	US-08-594-469-1	Sequence 1, Appli
15	2483.6	36.9	4410	US-08-906-957-1	Sequence 1, Appli
16	2439.8	36.3	5446	US-09-358-856C-1	Sequence 1, Appli
17	2188.4	32.5	7892	US-07-916-098A-40	Sequence 40, Appli
18	2185.6	32.5	7633	US-09-028-851-1	Sequence 1, Appli
19	2185.6	32.5	7633	US-08-815-520-1	Sequence 1, Appli
20	2185.6	32.5	7633	US-09-273-163-1	Sequence 1, Appli
21	2183.4	32.4	8068	US-08-809-513A-7	Sequence 7, Appli
22	2183.4	32.4	8068	US-09-301-593-27	Sequence 27, Appli
23	2183.4	32.4	8068	US-09-301-593-35	Sequence 35, Appli
24	2183	32.4	9144	US-08-556-978B-79	Sequence 79, Appli
25	2182.6	32.4	7731	US-09-301-593-29	Sequence 29, Appli
26	2182.6	32.4	7731	US-09-301-593-42	Sequence 42, Appli
27	2182	32.4	3699	US-08-053-131-120	Sequence 120, App

ALIGNMENTS

RESULT 1

US-09-183-188B-5
; Sequence 5, Application US/09183188B
; Patent No. 6491905
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.
; APPLICANT: Parker, William B.
; APPLICANT: Waud, William
; APPLICANT: Gadi, Vijayakrishna K.
; TITLE OF INVENTION: RECOMBINANT E. COLI FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12405/22
; CURRENT APPLICATION NUMBER: US/09/183,188B
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 08/702,181
; PRIOR FILING DATE: 1996-08-23
; PRIOR APPLICATION NUMBER: 08/122,321
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 60/064,676
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-183-188B-5

Query Match	57.3%	Score 3853.4;	DB 4;	Length 5013;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 1;	Indels 0;
Matches 3854;	Conservative	0;		Gaps 0;
QY	2875	AGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCTGTATACAGATTAAATCAGAAC	2934	
Db	1159	AGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCTGTATACAGATTAAATCAGAAC	1218	
QY	2935	GCAGAAGCGGTCTGTATAAACAAGATTTTCCTGGCGGAGTAGCGGGTGTCCACCTG	2994	
Db	1219	GCAGAAGCGGTCTGTATAAACAAGATTTTCCTGGCGGAGTAGCGGGTGTCCACCTG	1278	
QY	2995	ACCCCATCCGAACCTCAGAACTGAAACCCGTAGCGCGAGTGTGGGGTCTCCCC	3054	
Db	1279	ACCCCATCCGAACCTCAGAACTGAAACCCGTAGCGCGAGTGTGGGGTCTCCCC	1338	
QY	3055	ATGCGAGTAGTGGGAACCTCCAGGCGATCAAAATAAACAAGAGCTCAGTCGAAAGACTGG	3114	
Db	1339	ATGCGAGTAGTGGGAACCTCCAGGCGATCAAAATAAACAAGAGCTCAGTCGAAAGACTGG	1398	
QY	3115	GCCTTTCGTTTATCTGTTGTTTGTTCGTTGTAACCGTCTCCTGAGTAGGACAAATCGCGG	3174	
Db	1399	GCCTTTCGTTTATCTGTTGTTTGTTCGTTGTAACCGTCTCCTGAGTAGGACAAATCGCGG	1458	

QY 3175 GGAGCGGATTTGAACGTTGCGAAGCAACGCGCCCGAGGCGGTGCGGGCAGGACGCGCGCCA 3234
 Db 1459 GGAGCGGATTTGAACGTTGCGAAGCAACGCGCCCGAGGCGGTGCGGGCAGGACGCGCGCCA 1518
 QY 3235 TAAACTGCGGCGCATCAAAATTAAGCAGAGGCCATCCTGACGGATGGCCCTTTTTCGCTTT 3294
 Db 1519 TAAACTGCGGCGCATCAAAATTAAGCAGAGGCCATCCTGACGGATGGCCCTTTTTCGCTTT 1578
 QY 3295 CTACAAAACCTCTTTTGTATTTTCTAAATACATTTCAAAATATGATCCGCTCATGAGAC 3354
 Db 1579 CTACAAAACCTCTTTTGTATTTTCTAAATACATTTCAAAATATGATCCGCTCATGAGAC 1638
 QY 3355 AATAACCTGATAAATGCTTCAATAATATGAAAAGGAAGATGATGATTTCAACATT 3414
 Db 1639 AATAACCTGATAAATGCTTCAATAATATGAAAAGGAAGATGATGATTTCAACATT 1698
 QY 3415 TCCGTGTCGCGCTTATTCCTCTTTTGGGCAATTTGCTTCCCTCTTCTGTTTTCCTCACCCAG 3474
 Db 1699 TCCGTGTCGCGCTTATTCCTCTTTTGGGCAATTTGCTTCCCTCTTCTGTTTTCCTCACCCAG 1758
 QY 3475 AAAACCTGCTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCG 3534
 Db 1759 AAAACCTGCTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCG 1818
 QY 3535 AACTGGATCTCAACAGCGGTAAAGTCTTGAAGATTTTGGCCCGGAAAGACGTTTTCCAA 3594
 Db 1819 AACTGGATCTCAACAGCGGTAAAGTCTTGAAGATTTTGGCCCGGAAAGACGTTTTCCAA 1878
 QY 3595 TGATGAGCATTTTAAAGTCTGCTATGTCGGCGGTATATCCCGTGTGACGCGCGGC 3654
 Db 1879 TGATGAGCATTTTAAAGTCTGCTATGTCGGCGGTATATCCCGTGTGACGCGCGGC 1938
 QY 3655 AAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCG 3714
 Db 1939 AAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCG 1998
 QY 3715 TCACAGAAAAGCATCTTACGATGCGCATGACAGTAAGAGATTAATGCAAGTGTGCGCATAA 3774
 Db 1999 TCACAGAAAAGCATCTTACGATGCGCATGACAGTAAGAGATTAATGCAAGTGTGCGCATAA 2058
 QY 3775 CCATGAGTGAATAACTCGCGCACTTTACTTCTGACAAACGATCGGAGCGCGAAGGAGC 3834
 Db 2059 CCATGAGTGAATAACTCGCGCACTTTACTTCTGACAAACGATCGGAGCGCGAAGGAGC 2118
 QY 3835 TAAACGCTTTTGTGCAACAACTGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGG 3894
 Db 2119 TAAACGCTTTTGTGCAACAACTGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGG 2178
 QY 3895 AGCTGAATGAAGCCATACAAACGACGAGCGTGACACCAACGATGCTGTAGCAATGGCAA 3954
 Db 2179 AGCTGAATGAAGCCATACAAACGACGAGCGTGACACCAACGATGCTGTAGCAATGGCAA 2238
 QY 3955 CAACGTTGCGCAAACTATTAACTGCGCACTTTACTTCTAGCTTCCCGGCAACAATTA 4014
 Db 2239 CAACGTTGCGCAAACTATTAACTGCGCACTTTACTTCTAGCTTCCCGGCAACAATTA 2298
 QY 4015 TAGACTGATGAGCGGATAAAGTTGACGACCACTTTCTGCGCTCGGCCCTTTCGCGCTG 4074
 Db 2299 TAGACTGATGAGCGGATAAAGTTGACGACCACTTTCTGCGCTCGGCCCTTTCGCGCTG 2358
 QY 4075 GCTGTTTATGCTGATAAATCTGAGCGGTGAGCGTGGGTCTCGCGGTATCATTTGAG 4134
 Db 2359 GCTGTTTATGCTGATAAATCTGAGCGGTGAGCGTGGGTCTCGCGGTATCATTTGAG 2418
 QY 4135 CACTGGGCGCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGAGTCAGG 4194
 Db 2419 CACTGGGCGCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGAGTCAGG 2478
 QY 4195 CAACTATGATGAACGAATAGACAGATCGCTGAGATAGGTGCTCTCACTGATTAAGCAT 4254
 Db 2479 CAACTATGATGAACGAATAGACAGATCGCTGAGATAGGTGCTCTCACTGATTAAGCAT 2538

QY 4255 GGTAACTGTCAGACCAAGTTTACTCATATATACATTTAGATTGATTTTAAAACTTCATTTT 4314
 Db 2539 GGTAACTGTCAGACCAAGTTTACTCATATATACATTTAGATTGATTTTAAAACTTCATTTT 2598
 QY 4315 AATTTAAAGAGATCTAGGTGAAGATCCTTTTGTGATTAATCTCATGACCAAAATCCCTTAAC 4374
 Db 2599 AATTTAAAGAGATCTAGGTGAAGATCCTTTTGTGATTAATCTCATGACCAAAATCCCTTAAC 2658
 QY 4375 GTGAGTTTTCGTTCCACTGAGCGTCAGACCCGCTAGAAAAGATCAAAAGATCTTCTTGAG 4434
 Db 2659 GTGAGTTTTCGTTCCACTGAGCGTCAGACCCGCTAGAAAAGATCAAAAGATCTTCTTGAG 2718
 QY 4435 ATCCTTTTTCGCGCGTAACTCTGCTGCTTGGCAAAACAAAAACCCAGCTACCGCGG 4494
 Db 2719 ATCCTTTTTCGCGCGTAACTCTGCTGCTTGGCAAAACAAAAACCCAGCTACCGCGG 2778
 QY 4495 TGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTTTCGAAAGTAACTGGCTTCAGCA 4554
 Db 2779 TGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTTTCGAAAGTAACTGGCTTCAGCA 2838
 QY 4555 GAGCGAGATACCAAAATATCTGCTCTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGA 4614
 Db 2839 GAGCGAGATACCAAAATATCTGCTCTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGA 2898
 QY 4615 ACTGTAGCACCGCTACATACCTCGCTCTGCTAATCTCTTACCAGTGGCTGCGCA 4674
 Db 2899 ACTGTAGCACCGCTACATACCTCGCTCTGCTAATCTCTTACCAGTGGCTGCGCA 2958
 QY 4675 GTGCGGATAAGTGTCTTACCGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC 4734
 Db 2959 GTGCGGATAAGTGTCTTACCGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC 3018
 QY 4735 AGCGGTGCGGTAAACGCGGGGTTCTGTGCAACAGCCGAGCTTGGAGGAAACGACTTACA 4794
 Db 3019 AGCGGTGCGGTAAACGCGGGGTTCTGTGCAACAGCCGAGCTTGGAGGAAACGACTTACA 3078
 QY 4795 CCGAAGTGAATACCTACAGCGTGTATGAGAAAGCCGCTTCCGAAAGGAGAA 4854
 Db 3079 CCGAAGTGAATACCTACAGCGTGTATGAGAAAGCCGCTTCCGAAAGGAGAA 3138
 QY 4855 AGGCGGACAGGATATCCGGTAAGCGCGGTCGGAAACAGGAGAGCGCACGAGGAGCTTC 4914
 Db 3139 AGGCGGACAGGATATCCGGTAAGCGCGGTCGGAAACAGGAGAGCGCACGAGGAGCTTC 3198
 QY 4915 CAGGGGAAACGCTGTGTATCTTTATAGTCTGTGCGGTTTCGCCACTCTGACTTGAGC 4974
 Db 3199 CAGGGGAAACGCTGTGTATCTTTATAGTCTGTGCGGTTTCGCCACTCTGACTTGAGC 3258
 QY 4975 GTCGATTTTGTGATGCTCGTCAGGGGCGGAGCCTATGAAAACCGCAGCAACGCGG 5034
 Db 3259 GTCGATTTTGTGATGCTCGTCAGGGGCGGAGCCTATGAAAACCGCAGCAACGCGG 3318
 QY 5035 CTTTTCACGCTTCTGCGCTTTTGTGCTTTTGTCTCACATGTTCTTTCTGCTGTTAT 5094
 Db 3319 CTTTTCACGCTTCTGCGCTTTTGTGCTTTTGTCTCACATGTTCTTTCTGCTGTTAT 3378
 QY 5095 CCCCTGATTTCTGTGATAAACGCTATTAACGCTTTTGTGAGTGTGATACCGCTCGCCGCA 5154
 Db 3379 CCCCTGATTTCTGTGATAAACGCTATTAACGCTTTTGTGAGTGTGATACCGCTCGCCGCA 3438
 QY 5155 GCGGAAACGACGCGCAGCGAGTCAAGTACGAGGAGCGGAGCGGCTGATGCGGT 5214
 Db 3439 GCGGAAACGACGCGCAGCGAGTCAAGTACGAGGAGCGGAGCGGCTGATGCGGT 3498
 QY 5215 ATTTTCTCTTTACGCATCTGTGCGGTATTTTTCACACCGCATATGTTGCTACTCTCAGTACAA 5274
 Db 3499 ATTTTCTCTTTACGCATCTGTGCGGTATTTTTCACACCGCATATGTTGCTACTCTCAGTACAA 3558
 QY 5275 TCTGCTGTGATGCGCGCATAGTTAAGCAGATATACCTCGCTATCGCTACGTCAGTGGT 5334
 Db 3559 TCTGCTGTGATGCGCGCATAGTTAAGCAGATATACCTCGCTATCGCTACGTCAGTGGT 3618
 QY 5335 CATGCTGCGCCCGGACACCCCGCAACACCGCTGACGCGCCTTGAACGGGCTTGTCTGCT 5394

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 15:42:19 ; Search time 3779 Seconds
(without alignments)
11542.622 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 6729
Sequence: 1 gtttagacgttatcatcga.....agtttagcggaattgatctg 6729

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7297361 seqs, 3241162794 residues
Total number of hits satisfying chosen parameters: 14594722

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US04_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US02_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US01_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US00_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US10K_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US10L_PUBCOMB.seq:*
- 25: /cgn2_6/ptodata/1/pubpna/US10M_PUBCOMB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US10N_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6729	100.0	6729	21	US-10-649-433C-1
2	3853.4	57.3	5013	14	US-10-198-034-5
3	3852.4	57.3	5013	15	US-10-198-228-5
C 4	3852.2	57.2	5926	13	US-10-024-809-3
C 5	3852.2	57.2	5926	13	US-10-024-809-4
6	3850.6	57.2	4176	16	US-10-244-142A-3
C 7	3344.8	49.7	5024	17	US-10-313-963A-52

C	8	3344.8	49.7	5448	21	US-10-671-995A-2	Sequence 2, Appli
C	9	3344.8	49.7	5454	21	US-10-671-995A-1	Sequence 1, Appli
	10	3341.2	49.7	6464	15	US-10-151-690-20	Sequence 20, Appli
	11	2230.2	48.0	5398	22	US-10-505-082-3	Sequence 3, Appli
	12	2877.4	42.8	5522	10	US-09-872-868-16	Sequence 16, Appli
	13	2877.4	42.8	5525	10	US-09-872-868-17	Sequence 17, Appli
	14	2877.4	42.8	6086	10	US-09-872-868-12	Sequence 12, Appli
	15	2877.4	42.8	6086	10	US-09-872-868-13	Sequence 13, Appli
	16	2877.4	42.8	6086	10	US-09-872-868-14	Sequence 14, Appli
	17	2877.4	42.8	6088	10	US-09-872-868-15	Sequence 15, Appli
	18	2863.2	42.6	5451	20	US-10-485-281-6	Sequence 6, Appli
	19	2858	42.5	4557	17	US-10-457-372-5	GENERAL INFORMA
	20	2858	42.5	17135	10	US-09-843-150-55	Sequence 55, Appli
	21	2858	42.5	17135	20	US-10-475-962-55	Sequence 55, Appli
	22	2757.2	41.0	4700	17	US-10-313-963A-54	Sequence 54, Appli
	23	2743.4	40.8	5772	20	US-10-131-998A-37	Sequence 37, Appli
	24	2743.4	40.8	5772	20	US-10-131-998A-38	Sequence 38, Appli
	25	2731	40.6	4145	20	US-10-699-088-163	Sequence 163, App
	26	2731	40.6	4145	21	US-10-699-114-163	Sequence 163, App
	27	2729	40.6	4357	10	US-09-795-872-10	Sequence 10, Appli
	28	2729	40.6	4357	19	US-10-662-824-10	Sequence 10, Appli
	29	2664.2	39.6	6886	18	US-10-466-642-20	Sequence 20, Appli
	30	2664.2	39.6	6886	18	US-10-466-717-20	Sequence 20, Appli
	31	2658.4	39.5	5797	17	US-10-316-194-167	Sequence 167, App
	32	2658.4	39.5	5797	21	US-10-732-180-203	Sequence 203, App
	33	2613.4	38.8	6806	10	US-09-908-943A-194	Sequence 194, App
	34	2613.4	38.8	6806	20	US-10-801-487-194	Sequence 194, App
	35	2613.4	38.8	6806	20	US-10-801-938-194	Sequence 194, App
	36	2613.4	38.8	6806	20	US-10-801-509-194	Sequence 194, App
	37	2613.4	38.8	6806	20	US-10-801-486-194	Sequence 194, App
	38	2613.4	38.8	6806	21	US-10-801-493-194	Sequence 194, App
	39	2603.2	38.7	4921	21	US-10-495-491-2	Sequence 2, Appli
	40	2603.2	38.7	4945	21	US-10-495-491-3	Sequence 3, Appli
	41	2600.4	38.6	4935	21	US-10-495-491-1	Sequence 1, Appli
	42	2600.4	38.6	4951	21	US-10-495-491-4	Sequence 4, Appli
	43	2464.2	36.6	4956	16	US-10-227-617A-5	Sequence 5, Appli
	44	2439.8	36.3	5446	17	US-10-393-269-1	Sequence 1, Appli
C	45	2369	35.2	7615	10	US-09-898-616A-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-10-649-433C-1
; Sequence 1, Application US/10649433C
; Publication No. US20050118726A1
; GENERAL INFORMATION:
; APPLICANT: Schultz, Jerome S.
; TITLE OF INVENTION: System and Method for Detecting Bioanalytes and Method for
; TITLE OF INVENTION: Producing a Bioanalyte Sensor
; FILE REFERENCE: 03-016
; CURRENT APPLICATION NUMBER: US/10/649,433C
; CURRENT FILING DATE: 2003-08-26
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 6729
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-649-433C-1

Query Match	100.0%	Score	6729;	DB	21;	Length	6729;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	6729;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	GTTTGACAGCTTATCATCGACGTGACCAATGCTTCTGCGCTCAGGCAGCCATC	60				
Db	1	GTTTGACAGCTTATCATCGACGTGACCAATGCTTCTGCGCTCAGGCAGCCATC	60				
Qy	61	CGAAGCTGTGTATGGCTGTGTCAGGTCTTAATCACTCATTAATTCGTGCTCAAGC	120				
Db	61	CGAAGCTGTGTATGGCTGTGTCAGGTCTTAATCACTCATTAATTCGTGCTCAAGC	120				

QY 121 GCATCCCGTCTCGAATAATGTTTTTTCGCCGCAACATCATAAACGGTTCTGCAAAATATTC 180
DB 121 GCATCCCGTCTCGAATAATGTTTTTTCGCCGCAACATCATAAACGGTTCTGCAAAATATTC 180
QY 181 TGAATGAGCTGTGACAAATTAATCATCCGCTCGTATTAATGTGTGGAAATGTGAGCGGA 240
DB 181 TGAATGAGCTGTGACAAATTAATCATCCGCTCGTATTAATGTGTGGAAATGTGAGCGGA 240
QY 241 TAACAATTTTCAACAGGAAAACAGCCGCTCGAGAAAACGGAAGCGGCACCTGCTCTTTAA 300
DB 241 TAACAATTTTCAACAGGAAAACAGCCGCTCGAGAAAACGGAAGCGGCACCTGCTCTTTAA 300
QY 301 CAAATTTATCAGCAATCTGTGTGGCACTCAGACGGAAATTAATCGAATTAATTTATTA 360
DB 301 CAAATTTATCAGCAATCTGTGTGGCACTCAGACGGAAATTAATCGAATTAATTTATTA 360
QY 361 AAAATTAAGAGGTATATTAATGATATCGAATTAATTAAGGAGGAATAAACAATGGTGAG 420
DB 361 AAAATTAAGAGGTATATTAATGATATCGAATTAATTAAGGAGGAATAAACAATGGTGAG 420
QY 421 CAAGGGCAGGAGCTGTTTCAACGGGGTGGTCCCATCTCGTTCGAGCTGGACGGCAGCT 480
DB 421 CAAGGGCAGGAGCTGTTTCAACGGGGTGGTCCCATCTCGTTCGAGCTGGACGGCAGCT 480
QY 481 AAAAGGCCACAAGTTTACGGTGTCCGGCAGGGCGAGGGCGATGCACTACGGCAAGCT 540
DB 481 AAAAGGCCACAAGTTTACGGTGTCCGGCAGGGCGAGGGCGATGCACTACGGCAAGCT 540
QY 541 GACCTGAAGTTTCACTGCAACCGCAAGCTGCGCGCTGCGCGTCCGCTGGCCACCTCGTGAC 600
DB 541 GACCTGAAGTTTCACTGCAACCGCAAGCTGCGCGCTGCGCGTCCGCTGGCCACCTCGTGAC 600
QY 601 CACCTTCGGCTACGGCTCGAGTGTCTCCGCGCTTACCCCGCAACACATGAAGCAGCAGA 660
DB 601 CACCTTCGGCTACGGCTCGAGTGTCTCCGCGCTTACCCCGCAACACATGAAGCAGCAGA 660
QY 661 CTCTTCAAGTCCGCAATGCGCGAAGGCTAAGTTCAGAGAGCGCAACATCTCTTCAAGGA 720
DB 661 CTCTTCAAGTCCGCAATGCGCGAAGGCTAAGTTCAGAGAGCGCAACATCTCTTCAAGGA 720
QY 721 CGAGGCAACTACAGAGCCGCGCGAGTGAAGTTCCAGGCGCACACCTGCTGAACCG 780
DB 721 CGAGGCAACTACAGAGCCGCGCGAGTGAAGTTCCAGGCGCACACCTGCTGAACCG 780
QY 781 CATCGAGCTGAAGGSCATCGACTTCAAGAGGAGCGGCAACATCTCTGCGGCAACAGCTGA 840
DB 781 CATCGAGCTGAAGGSCATCGACTTCAAGAGGAGCGGCAACATCTCTGCGGCAACAGCTGA 840
QY 841 GTACAACTACAGAGCCAAAGCTTATATCATGCGCGCAAGCAGAGAAACGGCATCAA 900
DB 841 GTACAACTACAGAGCCAAAGCTTATATCATGCGCGCAAGCAGAGAAACGGCATCAA 900
QY 901 GGTGAATTTCAAGATCCGCCCAACATCGAGAGCGGCGAGCTGCGAGCTCGCGCACTA 960
DB 901 GGTGAATTTCAAGATCCGCCCAACATCGAGAGCGGCGAGCTGCGAGCTCGCGCACTA 960
QY 961 CCAGAGAACACCCCATCGGCGAGCGGCCCGTGTCTGCTGCCGCAACACCACTACCTGAG 1020
DB 961 CCAGAGAACACCCCATCGGCGAGCGGCCCGTGTCTGCTGCCGCAACACCACTACCTGAG 1020
QY 1021 CTACAGTCCGCTGAGCAAGAACCCCAACAGAGAGCGGATCACTGCTGCTGCTGAG 1080
DB 1021 CTACAGTCCGCTGAGCAAGAACCCCAACAGAGAGCGGATCACTGCTGCTGCTGAG 1080
QY 1081 GTTGTGACCGCGCGGATCACTCTCGGATGAGACGAGCTGTACAAGACTAGTGCTGA 1140
DB 1081 GTTGTGACCGCGCGGATCACTCTCGGATGAGACGAGCTGTACAAGACTAGTGCTGA 1140
QY 1141 TACTCGCAATGGTGTAACAATCTAATAGTACGAGATAACTTTATGCTGTAGTGCAGAA 1200
DB 1141 TACTCGCAATGGTGTAACAATCTAATAGTACGAGATAACTTTATGCTGTAGTGCAGAA 1200

QY 1201 GGCTATTGACGAAGATCGGAAAGCCGCGCAGATGTTTTCAGCTGCTGATGAATGATTCTCA 1260
DB 1201 GGCTATTGACGAAGATCGGAAAGCCGCGCAGATGTTTTCAGCTGCTGATGAATGATTCTCA 1260
QY 1261 GAATGACAGTCCCAAGCAGAACGATCAGATCGATGATTTGCTGGCCAAAGGGGTGAAGGC 1320
DB 1261 GAATGACAGTCCCAAGCAGAACGATCAGATCGATGATTTGCTGGCCAAAGGGGTGAAGGC 1320
QY 1321 ACTGCCCATCAACTGGTTGACCCGCGCAGCTGCGGGTACCGTGATTTGAGAAAGCGCGTGG 1380
DB 1321 ACTGCCCATCAACTGGTTGACCCGCGCAGCTGCGGGTACCGTGATTTGAGAAAGCGCGTGG 1380
QY 1381 GCAAAAACGTCGCGTGGTTTCTTCAACAAAGAACCGTCTCTGTAAGCGCTGATAGCTA 1440
DB 1381 GCAAAAACGTCGCGTGGTTTCTTCAACAAAGAACCGTCTCTGTAAGCGCTGATAGCTA 1440
QY 1441 CGACAAAGCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 CGACAAAGCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 GATTGCTAAAACATGCGGCGCGAATCAGGGTTGGGATCTGAAACAAAGACGCTCAGATTCA 1560
DB 1501 GATTGCTAAAACATGCGGCGCGAATCAGGGTTGGGATCTGAAACAAAGACGCTCAGATTCA 1560
QY 1561 GTTCGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
DB 1561 GTTCGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1621 GATTAAAGAAATGGAACGATTAAGGATCAAAACCTGAAACAGTTTACGTTAGATACCGCAAT 1680
DB 1621 GATTAAAGAAATGGAACGATTAAGGATCAAAACCTGAAACAGTTTACGTTAGATACCGCAAT 1680
QY 1681 GTGGGACACCGCTCAGGCGAAAGATAAGATGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
DB 1681 GTGGGACACCGCTCAGGCGAAAGATAAGATGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1741 CAAAATCGAAGTGTATTCGCAACCAAGATGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 1800
DB 1741 CAAAATCGAAGTGTATTCGCAACCAAGATGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 1800
QY 1801 GAAAGCACACAAAGTCCAGCATTCGCGTGTGTTGGCGTTCGATGCGCTGCGCAGAACGCT 1860
DB 1801 GAAAGCACACAAAGTCCAGCATTCGCGTGTGTTGGCGTTCGATGCGCTGCGCAGAACGCT 1860
QY 1861 GGCCTGTGTAATCCGCTGCACTGGCGGCGACCGTACTGAAACGATGCTAAACACAGGC 1920
DB 1861 GGCCTGTGTAATCCGCTGCACTGGCGGCGACCGTACTGAAACGATGCTAAACACAGGC 1920
QY 1921 GAAAGCGACCTTTCATCTGCGGAAACCTGCGCGATGTTAAAGTGGCGGCTGATGGCAC 1980
DB 1921 GAAAGCGACCTTTCATCTGCGGAAACCTGCGCGATGTTAAAGTGGCGGCTGATGGCAC 1980
QY 1981 CAACTGGAAGATCGACAAACAAAGTGGTCCGCGTACCTTATGTTGGCGTAGATAAAGACAA 2040
DB 1981 CAACTGGAAGATCGACAAACAAAGTGGTCCGCGTACCTTATGTTGGCGTAGATAAAGACAA 2040
QY 2041 CCTGCTGTAATTCAGCAAGAAAGGTAACAGTAAAGGAGAGAACTTTTCACTGGAGTTGT 2100
DB 2041 CCTGCTGTAATTCAGCAAGAAAGGTAACAGTAAAGGAGAGAACTTTTCACTGGAGTTGT 2100
QY 2101 CCCAATTTCTGTTGAATTTAGATGTTGTTAAATGGGCAAAATTTTCTGTCAGTGGAGA 2160
DB 2101 CCCAATTTCTGTTGAATTTAGATGTTGTTAAATGGGCAAAATTTTCTGTCAGTGGAGA 2160
QY 2161 GGGTGAAGGTGATGCAACATACGGAAACCTTACCTTTAAATTTTATTTTGCATCTACTGGAAA 2220
DB 2161 GGGTGAAGGTGATGCAACATACGGAAACCTTACCTTTAAATTTTATTTTGCATCTACTGGAAA 2220
QY 2221 ACTACCTGTTCCATGGGCAACACTGTGCACTACTTTTCTTATGTTGTTTCAATGCTTTTC 2280
DB 2221 ACTACCTGTTCCATGGGCAACACTGTGCACTACTTTTCTTATGTTGTTTCAATGCTTTTC 2280
QY 2281 CCGTTATCCGGATCATATGAAACGGCATGACTTTTTTCAAGAGTGCATGCTGCCGAGGTTA 2340

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:19:33 ; Search time 201.5 Seconds
(without alignments)
6426.229 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 11920
Sequence: 1 gtttagacgttatcatcg.....agttagcgcgaattgatctg 6729

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cpn2_1/USPTO spool_p/US10649433/runat_02082005_164945_1532/app_query.fasta_1.6919
-DB=PR_QPWT=faetan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649433 @CN 1 1 332 @runat_02082005_164945_1532 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1765	14.8	360	1 RPCL	lactose operon rep
2	1765	14.8	360	4 S58116	lac repressor - sy
3	1754	14.7	360	2 F90678	lac operon transcr
4	1747	14.7	360	2 B85529	lac operon transcr
5	1579	13.2	332	1 JGSCG	D-galactose-binding
6	1579	13.2	332	2 B91009	galactose-binding
7	1579	13.2	332	2 D85853	galactose-binding
8	1537	12.9	332	1 S15554	D-galactose-binding
9	1510	12.7	332	2 AF0781	D-galactose-binding
10	1508	12.7	332	1 T51301	galactose-binding
11	1467	12.3	286	2 T51301	beta-lactamase (EC
12	1467	12.3	286	2 S47061	beta-lactamase (EC
13	1467	12.3	286	4 S41975	beta-lactamase (EC
14	1464	12.3	286	1 PNECP	beta-lactamase (EC

15	1464	12.3	286	4 I40905	beta-lactamase (EC
16	1451	12.2	286	2 S60312	extended spectrum
17	1450	12.2	286	2 S60313	beta-lactamase (EC
18	1444	12.1	286	2 S60310	extended spectrum
19	1442	12.1	286	2 S60311	beta-lactamase (EC
20	1438	12.1	286	2 JQ1546	Bla protein - Salm
21	1425.5	12.0	335	2 AG0183	galactose-binding
22	1242	10.4	238	1 JQ1514	green-fluorescent
23	1167.5	9.8	324	2 C82214	galactoside ABC tr
24	1137.5	9.5	349	1 G64096	D-galactose-binding
25	960	8.1	286	2 S16146	beta-lactamase (EC
26	960	8.1	286	2 A60679	beta-lactamase (EC
27	956	8.0	286	1 A44998	beta-lactamase (EC
28	953	8.0	286	2 A37200	beta-lactamase (EC
29	952	8.0	286	2 A60632	beta-lactamase (EC
30	937	7.9	265	2 S04464	beta-lactamase (EC
31	931	7.8	265	2 S04334	beta-lactamase (EC
32	927.5	7.8	287	1 A44996	beta-lactamase (EC
33	924	7.8	265	2 A60448	beta-lactamase (EC
34	917	7.7	279	2 A24469	beta-lactamase (EC
35	896	7.5	286	1 A44958	beta-lactamase (EC
36	715	6.0	357	2 AE0104	lactose operon rep
37	674	5.6	149	2 JQ1541	hypothetical 16.9K
38	669.5	5.6	298	2 A41381	beta-lactamase (EC
39	632	5.3	354	2 B24925	lac repressor - K1
40	540.5	4.5	304	2 A35001	beta-lactamase (EC
41	536.5	4.5	304	2 A49789	beta-lactamase (EC
42	530.5	4.5	288	2 JQ1136	beta-lactamase (EC
43	524	4.4	105	2 JC2566	bla protein - pseu
44	519.5	4.4	281	2 D95395	probable Beta lact
45	518.5	4.3	288	2 JS0755	beta-lactamase (EC

ALIGNMENTS

RESULT 1

RPECL
lactose operon repressor - Escherichia coli (strain K-12)

N:Alternate names: lac repressor

C:Species: Escherichia coli

C:Date: 30-Apr-1982 #sequence revision 30-Apr-1982 #text change 09-Jul-2004

C:Accession: A93198; A91234; A92122; A93785; S40661; S02540; A64762; S68009; S14614; A03

R:Farabaugh, P.J.

Nature 274, 765-769, 1978

A:Title: Sequence of the lacI gene.

A:Reference number: A93198; MUID:78246991; PMID:355891

A:Accession: A93198

A:Molecule type: DNA

A:Residues: 1-360 <FAR>

A:Cross-references: UNIPROT:P03023

R:Beyreuther, K.; Adler, K.; Fanning, E.; Murray, A.; Geisler, N.

Eur. J. Biochem. 59, 491-509, 1975

A:Title: Amino-acid sequence of lac repressor from Escherichia coli. Isolation, sequence

A:Reference number: A91234; MUID:76091932; PMID:1107032

A:Accession: A91234

A:Molecule type: protein

A:Residues: 1-147;159-163,'O',165-230;233-360 <BEY>

A:Note: the active repressor is a tetramer of identical chains

A:Note: this protein was obtained from a strain with the I-SQ mutation, which leads to a

A:Note: this sequence has since been revised

R:Beyreuther, K.

Nature 274, 767, 1978

A:Reference number: A93199

A:Contents: annotation; revision

A:Note: the revised sequence is identical with that shown

R:Platt, T.; Files, J.G.; Weber, K.

J. Biol. Chem. 248, 110-121, 1973

A:Reference number: A92122; MUID:73143730; PMID:4571224

A:Accession: A92122

A:Molecule type: protein

A:Residues: 1-59;96-101;206-215;328-347 <PLA>

A:Note: removal of residues 1-59 and 328-347 results in the molecule losing its DNA-bin

R:Ganem, D.; Miller, J.H.; Files, J.G.; Platt, T.; Weber, K.

Proc. Natl. Acad. Sci. U.S.A. 70, 3165-3169, 1973
A>Title: Reinitiation of a lac repressor fragment at a codon other than ATG.
A;Reference number: A93785; MUID:74126378; PMID:4594037
A;Accession: A93785
A;Molecule type: protein
A;Residues: 60-70;73-78;83-86 <GN>
R;Shin, J.A.; Ebricht, R.H.; Dervan, P.B.
Nucleic Acids Res. 19, 5233-5236, 1991
A>Title: Orientation of the lac repressor DNA binding domain in complex with the left la
A;Reference number: S40661; MUID:92020210; PMID:1923807
A;Accession: S40661
A;Molecule type: DNA
A;Residues: 1-56 <SHI>
R;Gordon, A.J.E.; Burns, P.A.; Fix, D.P.; Yatagai, F.; Allen, F.L.; Horsfall, M.J.; Hall
J. Mol. Biol. 200, 239-251, 1988
A>Title: Missense mutation in the lacI gene of *Escherichia coli*. Inferences on the struc
A;Reference number: S02540; MUID:88230449; PMID:3286877
A;Accession: S02540
A>Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-60 <GOR>
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of *Escherichia coli* K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64720
A>Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-360 <BLAT>
A;Cross-references: GB:A000141; GB:U00096; MUID:91786532; PIDN:AAC73448.1; PID:91786540;
A;Experimental source: strain K-12, substrain MG1655
R;Khanashev, D.E.; Esipova, N.G.; Ebraldise, K.K.; Mirzabekov, A.D.
FEBS Lett. 375, 27-30, 1995
A>Title: Mechanism of Lac repressor switch-off: orientation of the Lac repressor DNA-bin
A;Reference number: S68009; MUID:96087076; PMID:7498473
A;Accession: S68009
A;Molecule type: protein
A;Residues: 'X', 2-22; '32-32', 'X', 34-35 <XAM>
R;Chen, J.; Matthews, K.K.S.M.
submitted to the EMBL Data Library, March 1991
A;Description: T41 mutation in lac repressor is Tyr282Asp.
A;Reference number: S14614
A;Accession: S14614
A;Molecule type: DNA
A;Residues: 'V', 2-108, 'T', 110-281, 'D', 283-285, 'S', 287-360 <CHE>
A;Cross-references: EMBL:X58469
A;Experimental source: T41 mutant
C;Genetics:
A;Gene: lacI
A;Map position: 8 min
A;Start codon: GTG
C;Complex: homotetramer
C;Superfamily: lac repressor
C;Keywords: DNA binding; homotetramer; transcription regulation
F;5-25/Region: helix-turn-helix motif

Alignment Scores:
Pred. No.: 2, 07e-100 Length: 360
Score: 1765.00 Matches: 358
Percent Similarity: 99.72% Conservative: 1
Best Local Similarity: 99.44% Mismatches: 1
Query Match: 14.81% Indels: 0
DB: 1 Gaps: 0

US-10-649-433C-1 (1-6729) x RPECL (1-360)

QY 5608 GTGAACAGTAACGTTTACGATGCCAGAGTATGCCGGTGCTCTTATCAGACCGTT 5667
Db 1 MetLysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrVal 20

QY 5668 TCCCGCGTGGTGAACAGCCAGCCAGCCGTCCTTCTGCGAAAAACGCGGGAAGAGTGAAGCG 5727
Db 1 MetLysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrVal 20

RESULT 2
S58116
lac repressor - synthetic
C;Species: synthetic
C;Date: 13-Jan-1996 #sequence_revision 22-Aug-1996 #text_change 20-Oct-2000
C;Accession: S58116
R;Kahrs, A.F.; Odenbreit, S.; Schmitt, W.; Heuermann, D.; Meyer, T.F.; Haas, R.

Db 21 SerArgValValAsnGlnAlaSerHisValSerAlaLysThrArgGluLysValGluAla 40
QY 5728 GCGATGGCGGAGCTGAATACATTCCCAACCGCTGGCACAACAACCTGGCGGCAACAG 5787
Db 41 AlaMetAlaGluLeuAenTyrIleProAsnArgValAlaGlnLeuAlaGlyLysGln 60
QY 5788 TCGTGTCTGATTGGCGTTCACCTCCAGTCTGCCCTGCACGCGCGCTGCCAAATGTC 5847
Db 61 SerLeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSerGlnIleVal 80
QY 5848 GCGGCGATTAAATCTCGCGCGATCAACTGGTGGTCCAGCGTGGTGGTCTCGATGTAGAA 5907
Db 81 AlalaleIysSerArgAlaAspGlnLeuGlyAlaSerValValSerMetValGlu 100
QY 5908 CGAAGCGCGTCGAAGCCTGTAAAGCGCGGTGCACAATCTTCTCGGCAACCGTCAGT 5967
Db 101 ArgSerGlyValGluAlaCysLysAlaValHisAsnLeuLeuAlaGlnArgValSer 120
QY 5968 GGGCTGATCATTAATCTCCGCTCGATGCCAGATGCCATTGCTGTGGAGAGTCCTGC 6027
Db 121 GlyLeuIleIleAsnTyrProLeuAspGlnAspAlaIleAlaValGluAlaCys 140
QY 6028 ACTAATGTTCGGCGTATTCTTGTATGCTCTCCAGCAGACACCATCAACAGATTATT 6087
Db 141 ThrAsnValProAlaLeuPheLeuAspValSerAspGlnThrProIleAsnSerIleIle 160
QY 6088 TTCTCCCATGAAGACGGTAGCGAGCTGGCGGTGGAGCATCTGGTCCGATTCGGTCA 6147
Db 161 PheSerHisGluAspGlyThrArgLeuGlyValGluHisLeuValAlaLeuGlyHisGln 180
QY 6148 CAATCGCGCTGTAGCGGCGCCATTAAAGTCTCTCTCGGCGCTGCTGCTGCTGCTGC 6207
Db 181 GlnIleAlaLeuLeuAlaGlyProLeuSerSerValSerAlaArgLeuArgLeuAlaGly 200
QY 6208 TGGCATAAATATCTCACTCGCAATCAATTCAGCCGATAGCGAAGCAAGGAGCGACTGG 6267
Db 201 TrpHisIysTyrLeuThrArgAsnGlnIleGlnProIleAlaGluArgGluGlyAspTrp 220
QY 6268 AGTCCCATGTCCGGTTTCAACAAACCATGCAATGTAATGAGGGCATCGTTCCCACT 6327
Db 221 SerAlaMetSerGlyPheGlnGlnThrMetGlnMetLeuAsnGluGlyIleValProThr 240
QY 6328 GCGATGTCTGTTGCCAAGCATCAGATGGCGTGGCGCAATGGCGCATTTACCGAGTCC 6387
Db 241 AlaMetLeuValAlaAsnAspGlnMetAlaLeuGlyAlaMetArgAlaIleThrGluSer 260
QY 6388 GGGCTGGCGTGTGGTGGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCA 6447
Db 261 GlyLeuArgValGlyAlaAspIleSerValValGlyTyrAspAspThrGluAspSerSer 280
QY 6448 TGTATATCCCGCGTCAACCCATCAACAGAGATTTTCGCTGTGGGGCAACCAAGC 6507
Db 281 CysTyrIleProLeuThrThrIleLysGlnAspPheArgLeuLeuGlyGlnThrSer 300
QY 6508 GTGACCGCTGTGCAACTCTCTCAGGCGCGCGGTGAGGGCAATCAGCTGTGCGCC 6567
Db 301 ValAspArgLeuLeuLeuSerGlnGlyAlaValLysGlyAsnGlnLeuLeuPro 320
QY 6568 GTCTCACTGGTGAAGAAAAACCAACCCCTGGCGCCCAATAGCAACCGCTCTCCCGC 6627
Db 321 ValSerLeuValLysArgLysThrThrLeuAlaProAsnThrGlnThrAlaSerProArg 340
QY 6628 GCGTTGGCCGATTCATTATGACGTGGCAGCAGAGTTCCTCCGACTGGAAAGCGGCGAG 6687
Db 341 AlaLeuAlaAspSerLeuMetGlnLeuAlaArgGlnValSerArgLeuLeuGlyGln 360

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 01:59:21 ; Search time 19523 Seconds

(without alignments)
13119.627 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttgacgcttattatcg.....agttagcggaattgatctg 6729

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 1903213700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gse1:*
9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1055.8	15.7	1070	1	AJ281552 4A3A-P6F1
C 2	1018.2	15.1	1048	7	CO552396 ACLV4_50
C 3	984.8	14.6	1013	4	BM438846 IplvY0015
C 4	924	13.7	1004	1	AJ281480 4A3A-P4G8
C 5	889.8	13.2	928	7	CO487414 GQ0227.B7
C 6	845	12.6	902	7	CR753463 DKFZp469G
C 7	837.8	12.5	841	1	AL042026 DKFZp434E
C 8	835.8	12.4	935	4	BG838279 Gc01_10e0
C 9	823	12.2	854	4	BM438950 IplvY0049
C 10	820.2	12.2	865	7	CK125894 BES182411
C 11	812.8	12.1	1126	8	BZ577702 msh2_5533
C 12	806.4	12.0	819	6	CD649375 CvGnd0008
C 13	800.6	11.9	856	7	CN823189 Oa splbn
C 14	796	11.8	1011	8	BZ576726 msh2_5071
C 15	791.4	11.8	827	7	CN823302 Oa splbn
C 16	785.4	11.7	1574	8	BZ572566 msh2_2693
C 17	779.4	11.6	1073	7	CF269652 Fcylcoid8
C 18	778	11.6	794	5	BQ751936 ESTG32499
C 19	777.8	11.6	800	1	AJ281449 4A3A-P4D5
C 20	773.2	11.5	1336	8	BZ575810 msh2_4637
C 21	772.8	11.5	789	6	CD280920 G44224.42
C 22	772	11.5	846	7	CV468077 est_1_van
C 23	770.6	11.5	779	7	CK394523 hggadaD0
C 24	770.4	11.4	1249	8	BZ572284 msh2_2572

ALIGNMENTS

RESULT 1	AJ281552	1070 bp	mRNA	linear	EST 30-JUN-2000	
LOCUS	4A3A-P6F11-F	Anopheles gambiae	immune competent	4A3A	Anopheles	
DEFINITION	gambiae cDNA clone 4A3A-P6F11, mRNA sequence.					
ACCESSION	AJ281552					
VERSION	AJ281552.1	GI:6929432				
KEYWORDS	EST.					
SOURCE	Anopheles gambiae	(African malaria mosquito)				
ORGANISM	Anopheles gambiae					
REFERENCE	1 (bases 1 to 1070)					
AUTHORS	Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C., Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B. and Kafatos,F.C.					
TITLE	Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines					
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (12)	6619-6624	(2000)		
MEDLINE	20300950					
PUBMED	10841561					
COMMENT	Contact: Dimopoulos G Fotis C. Kafatos laboratory European Molecular Biology Laboratory Meyerhofstrasse 1, 69117 Heidelberg, Germany.					
FEATURES	source					
	1..1070					
	/organism="Anopheles gambiae"					
	/mol_type="mRNA"					
	/strain="4A r/r"					
	/db_xref="taxon:7165"					
	/clone="4A3A-P6F11"					
	/cell_line="immune competent 4A3A"					
	/lab_host="E. coli DH10B"					
	/clone_lib="Anopheles gambiae immune competent 4A3A"					
	/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI; Site 2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996) : Normalization and Subtraction: Two approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."					
ORIGIN						

25	769.2	11.4	832	7	CN822433	Oa splbn
26	766.4	11.4	769	7	CK118014	218n06.pl
27	764.2	11.4	769	7	CV224987	CS hyp_24
C 28	763.8	11.4	966	8	BZ570738	msh2_1513
C 29	762.4	11.3	793	7	CK122681	BES182410
30	762	11.3	793	5	BQ751655	ESTG32218
31	761.6	11.3	780	5	BQ825693	103012980
32	758.2	11.3	820	7	CR753457	DKFZp469P
33	755.4	11.2	759	6	CD279661	G43818_35
34	755.4	11.2	806	5	BQ751220	ESTG31783
35	753.8	11.2	1463	8	BZ571475	msh2_1906
36	749.8	11.1	786	7	CN823164	Oa splbn
37	747.4	11.1	752	7	CR766850	DKFZp468H
C 38	745.4	11.1	844	8	BZ574513	msh2_3706
C 39	742	11.0	832	4	BG923768	602825893
C 40	741.4	11.0	758	9	CL422788	AE0544_Sa
41	739.4	11.0	741	6	CD279174	G44221_83
C 42	739.2	11.0	914	8	BZ569398	pacc2-164
43	733.2	10.9	811	9	ATH517156	AJ517156 Arabidops
44	733.2	10.9	1370	8	BZ571721	msh2_2025
45	732.2	10.9	950	8	BZ571129	msh2_1741

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:11:07 ; Search time 1027.5 Seconds
(without alignments)
6707.116 Million cell updates/sec

Title: US-10-649-433c-1
Perfect score: 11920
Sequence: 1 gtttagacgttatcatcg.....agtttagcgcgaattgatctg 6729

Scoring table:

BGAPop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cn2.1/USPTO spool_P/US10649433/runat_02082005_164945_1521/app.query.fasta_1.6919
-DB=UniProt -QFWT=fascan -SUFFIX=rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CN 1.1 1271 @runat_02082005_164945_1521 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1765	14.8	360	1 LACI_ECOLI
2	1754	14.7	360	2 Q7AH57
3	1747	14.7	360	2 Q8X684
4	1579	13.2	332	1 DGAL_ECOLI
5	1579	13.2	332	2 Q8X641
6	1537	12.9	332	1 DGAL_CITPR
7	1518	12.7	332	1 DGAL_SALTY
8	1510	12.7	332	2 Q8Z5A4
9	1467	12.3	286	2 Q38058
10	1467	12.3	286	2 Q00626
11	1467	12.3	286	2 Q79CL6
12	1467	12.3	286	2 Q79DR3
13	1464	12.3	286	1 BLAT_ECOLI
14	1464	12.3	286	1 BLAT_SALTY
15	1464	12.3	286	2 Q6A253
16	1464	12.3	286	2 Q61BN9

17	1464	12.3	286	2 Q6LCV6
18	1464	12.3	286	2 Q6TMH1
19	1464	12.3	286	2 Q6W9J1
20	1464	12.3	286	2 Q6WRX2
21	1464	12.3	286	2 Q6WZD4
22	1464	12.3	286	2 Q7B3X5
23	1464	12.3	286	2 Q7B899
24	1464	12.3	286	2 Q7BP57
25	1464	12.3	286	2 Q7BR75
26	1464	12.3	286	2 Q7DFY3
27	1464	12.3	286	2 Q7DHD3
28	1464	12.3	286	2 Q799Y1
29	1464	12.3	286	2 Q7BVP8
30	1461	12.3	286	2 Q8KSD2
31	1460	12.2	286	2 Q8KSD3
32	1460	12.2	286	2 Q8KSD3
33	1460	12.2	286	2 Q932Y6
34	1460	12.2	286	2 Q93A80
35	1460	12.2	286	2 Q6QRI1
36	1460	12.2	286	2 Q6SQJ9
37	1460	12.2	286	2 Q6UK84
38	1460	12.2	286	2 Q6UVM7
39	1460	12.2	286	2 Q6W7J4
40	1460	12.2	286	2 Q7B0V0
41	1460	12.2	286	2 Q9RMS2
42	1460	12.2	286	2 Q9RN48
43	1459	12.2	286	2 Q93A77
44	1459	12.2	286	2 Q9R745
45	1458	12.2	286	2 Q48406

ALIGNMENTS

RESULT 1
LACI_ECOLI STANDARD; PRT; 360 AA.
ID LACI_ECOLI
AC P03023; O09196; P71309; Q47338;
DT 21-JUL-1986 (Rel. 01, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Lactose operon repressor.
GN Name=laci; OrderedLocusNames=b0345;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=78246991; PubMed=355891;
RA Farabaugh P.J.;
RT "Sequence of the lacI gene."
RL Nature 274:765-769(1978).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [16]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [17]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [18]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [19]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [25]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [26]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [27]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [28]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [29]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [30]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [31]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [32]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [33]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [34]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [35]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [36]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [37]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [38]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F

RP SEQUENCE FROM N.A.
RA Marsh S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 1-147; 159-230 AND 233-360.
RX MEDLINE=76091932; PubMed=1107032;
RA Beyreuther K., Adler K., Fanning E., Murray C., Klemm A., Geisler N.;
RT "Amino-acid sequence of lac repressor headpiece from *Escherichia coli*.
RT Isolation, sequence analysis and sequence assembly of tryptic peptides
RT and cyanogen-bromide fragments.";
RL Eur. J. Biochem. 59:491-509(1975).
[6]
RP SEQUENCE OF 1-59; 96-101; 206-215 AND 328-347.
RX MEDLINE=73143730; PubMed=4571224;
RA Platt T., Files J.G., Weber K.;
RT "Lac repressor. Specific proteolytic destruction of the NH 2 -terminal
RT region and loss of the deoxyribonucleic acid-binding activity.";
RL J. Biol. Chem. 248:110-121(1973).
[7]
RP SEQUENCE OF 60-70; 73-78 AND 83-86.
RX MEDLINE=74126378; PubMed=4594037;
RA Ganem D., Miller J.H., Files J.G., Platt T., Weber K.;
RT "Reinitiation of a lac repressor fragment at a condon other than
RT AUG.";
RL Proc. Natl. Acad. Sci. U.S.A. 70:3165-3169(1973).
[8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=88230449; PubMed=3286877;
RA Gordon A.J.E., Burns P.A., Fix D.F., Yatagai F., Allen F.L.,
RA Horsfall M.J., Halliday J.A., Gray J., Bernlot-Moens C.,
RA Gluckman B.W.;
RT "Missense mutation in the lacI gene of *Escherichia coli*. Inferences on
RT the structure of the repressor protein.";
RL J. Mol. Biol. 200:239-251(1988).
[9]
RP MUTAGENESIS.
RX MEDLINE=90189956; PubMed=2178920;
RA Lehming N., Sartorius J., Kisters-Woike B., von Wilcken-Bergmann B.,
RA Mueller-Hill B.;
RT "Mutant lac repressors with new specificities hint at rules for
RT protein-DNA recognition.";
RL EMBO J. 9:615-621(1990).
[10]
RP MUTAGENESIS.
RX MEDLINE=94322386; PubMed=8046748;
RA Markiewicz P., Kleina L.G., Cruz C., Ehret S., Miller J.H.;
RT "Genetic studies of the lac repressor. XIV. Analysis of 4000 altered
RT *Escherichia coli* lac repressors reveals essential and non-essential
RT residues, as well as 'spacers' which do not require a specific
RT sequence.";
RL J. Mol. Biol. 240:421-433(1994).
[11]
RP 3D-STRUCTURE MODELING.
RX MEDLINE=91249837; PubMed=2040302;
RA Kisters-Woike B., Lehming N., Sartorius J., von Wilcken-Bergmann B.,
RA Mueller-Hill B.;
RT "A model of the lac repressor-operator complex based on physical and
RT genetic data.";
RL Eur. J. Biochem. 198:411-419(1991).
[12]
RP STRUCTURE BY NMR.
RX MEDLINE=89113344; PubMed=3064080;
RA Boelens R., Lamerichs R.M.J.N., Rullmann J.A.C., van Boom J.H.,
RA Kaptein R.;
RT "The interaction of lac repressor headpiece with its operator: an NMR
RT view.";
RL Protein Seq. Data Anal. 1:487-498(1988).
[13]
RP STRUCTURE BY NMR.
RX MEDLINE=89302886; PubMed=2742823;
RA Lamerichs R.M.J.N., Boelens R., van der Marel G.A., van Boom J.H.,
RA Kaptein R., Buck F., Fera B., Rueterjans H.;
RT "H NMR study of a complex between the lac repressor headpiece and a 22
RT base pair symmetric lac operator.";
RL Biochemistry 28:2985-2991(1989).
[14]
RP STRUCTURE BY NMR OF 1-56.
RX MEDLINE=96275660; PubMed=8683581; DOI=10.1006/jmbi.1996.0356;
RA Sljaper M., Bonvin A.M., Boelens R., Kaptein R.;
RT "Refined structure of lac repressor headpiece (1-56) determined by
RT relaxation matrix calculations from 2D and 3D NOE data: change of
RT tertiary structure upon binding to the lac operator.";
RL J. Mol. Biol. 259:761-773(1996).
[15]
RP STRUCTURE BY NMR OF 1-62.
RX MEDLINE=20113476; PubMed=10647179; DOI=10.1016/S0969-2126(00)88339-2;
RA Spronk C.A., Bonvin A.M., Radha P.K., Melacini G., Boelens R.,
RA Kaptein R.;
RT "The solution structure of Lac repressor headpiece 62 complexed to a
RT symmetrical lac operator.";
RL Structure 7:1483-1492(1999).
[16]
RP X-RAY CRYSTALLOGRAPHY (4.8 ANGSTROMS).
RX MEDLINE=96239623; PubMed=8638105;
RA Lewis M., Chang G., Horton N.C., Kercher M.A., Pace H.C.,
RA Schumacher M.A., Brennan R.G., Lu P.;
RT "Crystal structure of the lactose operon repressor and its complexes
RT with DNA and inducer.";
RL Science 271:1247-1254(1996).
[17]
CC -!- FUNCTION: Repressor of the lactose operon. Binds lactose as an
CC inducer.
CC -!- SUBUNIT: Homotetramer.
CC -!- MISCELLANEOUS: Removal of residues 1-59 and 328-347 results in the
CC molecule losing its DNA-binding activity while maintaining its
CC inducer-binding activity and its tetrameric structure.
CC -!- SIMILARITY: Contains 1 HTH lacI-type DNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V00294; CAA23569.1; -;
CC EMBL; J01636; AAA24052.1; -;
CC EMBL; U00096; AAC73448.1; -;
CC EMBL; U73857; AAB18069.1; ALT_INIT.
CC EMBL; X58469; CAA41383.1; -;
CC EMBL; U86347; AAB47270.1; ALT_INIT.
CC EMBL; U72488; AAB36549.1; -;
CC EMBL; U78872; AAB37348.1; -;
CC EMBL; U78873; AAB37351.1; -;
CC EMBL; U78874; AAB37354.1; -;
CC PIR; A93198; RPECL.
CC PDB; 1CJG; NMR; A/B=1-62.
CC PDB; 1EFA; X-ray; A/B/C=1-333.
CC PDB; 1JWL; X-ray; A/B/C=1-333.
CC PDB; 1JYE; X-ray; A=1-349.
CC PDB; 1JYF; X-ray; A=1-349.
CC PDB; 1L1M; NMR; A/B=1-62.
CC PDB; 1LBG; X-ray; A/B/C/D=1-359.
CC PDB; 1LBH; X-ray; A/B/C/D=1-359.
CC PDB; 1LBI; X-ray; A/B/C/D=1-359.
CC PDB; 1LCC; NMR; A=1-51.
CC PDB; 1LGD; NMR; A=1-51.
CC PDB; 1LQC; NMR; @=1-56.
CC PDB; 1LTP; Model; l=1-262.
CC PDB; 1TLF; X-ray; A/B/C/D=60-360.
CC ECO2DBASE; H039.0; 6TH EDITION.
CC EcoGene; EG10525; lacI.
CC InterPro; IPR000843; HTH LacI.
CC InterPro; IPR010982; Lambda_like_DNA.